TECHNISCHE
UNIVERSITÄT
DARMSTADT

Mobile Interaction with Large Multimedia Information Spaces

Dissertation
Zur Erlangung des akademischen Grades
eines Doktor rerum naturalium (Dr. rer. nat.)

Eingereicht von
Dipl.-Inform., Dipl.-Math. Jochen Huber
geb. in Worms

Angenommen vom
Fachbereich Informatik der Technischen Universität Darmstadt

Gutachter: Prof. Dr. Max Mühlhäuser (Technische Universität Darmstadt, DE)
Prof. Lawrence A. Rowe, Ph.D. (FX Palo Alto Laboratory Inc., USA)

Tag der Einreichung: 29. Oktober 2012
Tag der Disputation: 17. Dezember 2012

D17–Darmstadt 2012

Abstract

The proliferation of mobile devices such as the Apple iPhone entails ubiquitous availability of a variety of multimedia information to users on the move; ready at our fingertips, virtually *anywhere* and *anytime*. However, these devices have strong limitations: small screen real estate and restricted input capabilities. Moreover, with the ever increasing number of multimedia data and services, the navigation complexity rises and in turn decreases both user experience and usability. This thesis explores how the affordances of mobile devices can be leveraged to design novel interfaces for the mobile interaction with large multimedia information spaces, toward a more usable and enjoyable user experience. The contributions are placed alongside three main research directions:

The first, *device-centric interaction*, investigates how mobile devices can be used efficiently in spite of the small screen leveraging input modalities such as touch or tilt. It focuses on mobile video browsing as a guiding scenario and contributes an exploration of the design space of mobile video browsers. It further contributes a set of novel interfaces, which are analyzed in an extensive controlled experiment. To state only one example, our analysis shows that the video player included in the iPhone has significant drawbacks; it also points out how future versions can be designed to be more usable.

The second research direction, *space-centric interaction*, pushes the boundaries of the virtual information space beyond the small screen toward the physical environment. It particularly looks at how mobile devices can be used as see-through displays when mapping the virtual information to the physical space. It also contributes an empirically grounded theory thereof and a novel movement time model for the embodied navigation with spatially-aware displays.

The third axis of research is devoted to the questions of how virtual and physical space can be tightly integrated, allowing for *both device- and space-centric interaction* with real-world objects–not just one particular mobile device. Pico projectors are used to project virtual information into physical space, turning objects into dedicated projection surfaces and tangible interaction devices. The thesis investigates this in an exploratory field study. Based upon this study, it advances the field of pico projector interaction by deriving novel interaction techniques. Results from a user study indicate the potential to fundamentally change how we ubiquitously interact with augmented real-world objects.

Zusammenfassung

Leistungsfähige mobile Endgeräte wie etwa das iPhone von Apple ermöglichen es Benutzern überall und jederzeit auf eine Vielzahl von Multimedia-Informationen zuzugreifen. Sie besitzen jedoch wesentliche Einschränkungen: kleine Bildschirme und begrenzte Eingabemöglichkeiten. Zusätzlich erhöht die stetig wachsende Menge an Multimedia-Informationen die Komplexität der Interaktion. Insgesamt führt dies zu einer schlechten Benutzbarkeit und User Experience. Diese Dissertation exploriert, wie die Charakteristika mobiler Endgeräte für die Gestaltung benutzbarer Benutzerschnittstellen zur Interaktion mit großen multimedialen Informationsräumen ausgenutzt werden können. Die Beiträge sind in drei Forschungsrichtungen verankert:

Die erste Forschungsrichtung, *Device-Centric Interaction*, untersucht wie mobile Endgeräte trotz ihrer Einschränkungen effizient benutzt werden können. Hierbei dient die mobile Video-Navigation als Leitszenario. Die Beiträge umfassen eine Exploration des Gestaltungsraums mobiler Video-Browser, sowie neue Benutzerschnittstellen, die in einem umfangreichen kontrollierten Experiment analysiert wurden. Die Ergebnisse zeigen, dass beispielsweise der Video-Browser des iPhones gravierende Nachteile hat. Zudem werden Implikationen abgeleitet, wie zukünftige Schnittstellen benutzbarer gestaltet werden können.

Die zweite Forschungsrichtung, *Space-Centric Interaction*, zielt darauf ab die Bildschirmgrenzen der mobilen Geräte zu überwinden. Dabei werden mobile Endgeräte als sogenannte “See-through Displays” genutzt, um den virtuellen Informationsraum auf den physischen Raum abzubilden und darin zu verankern. Die Beiträge umfassen eine empirisch gestützte Theorie der Interaktion mit See-through Displays und ein theoretisches Modell, welches die Navigationszeit in Abhängigkeit von der räumlichen Bewegung des Displays modelliert.

Die dritte Forschungsrichtung integriert die Sichtweisen der *Device-* und *Space-Centric Interaction*: virtuelle Informationen werden mittels Pico Projektoren in den physischen Raum projiziert. Alltagsgegenstände werden so zu dedizierten Projektionsflächen und können zur begreifbaren Interaktion genutzt werden. Dies wird in einer explorativen Feldstudie untersucht. Aufbauend auf den Ergebnissen werden neuartige Interaktionstechniken vorgestellt. Deren Evaluationsergebnisse zeigen das Potenzial auf, die ubiquitäre Interaktion mit digital angereicherten Objekten fundamental zu verändern.

Acknowledgements

This thesis would not have been possible without the support of my advisors, fellow colleagues, students and friends.

In particular, I thank my advisor and mentor Max Mühlhäuser (Technische Universität Darmstadt, Germany). I am thankful for his great support, his passion for research and many fruitful and inspiring discussions–space and opportunities provided for my research are invaluable to me. Thank you, Max!

I am also very grateful to Lawrence A. Rowe (FX Palo Alto Laboratory, USA) for acting as co-referee and having me as a visiting student at FXPAL in Spring 2011. My stay allowed me to conduct research substantial to this thesis for which I am thankful. On the same note, I would like to thank Chunyuan Liao and Qiong Liu for their support and advice during my stay.

Furthermore, I wish to extend special thanks to Jürgen Steimle (MIT Media Lab, USA and Max Planck Institute for Informatics, Germany) for his continuous support, advice and mentorship during my dissertation work.

I am very grateful for having been a member of the research training group on "e-learning" funded by the German Research Foundation. In particular, I would like to thank Andreas Kaminski and Werner Sesink for shaping my view upon inter-disciplinary research and the many inspiring discussions we had.

I am immensely thankful for all the support I received from present and former members of the Telecooperation Lab at Technische Universität Darmstadt. I wish to particularly thank Markus Weimer, Thorsten Strufe, Daniel Schreiber, Felix Heinrichs, Simon Olberding, Melanie Hartmann and Elke Halla. I would also like to thank all of my students who contributed to my research projects: Sascha Bleidner, Fawaz Amjad Malik, Markus Müller, Faheem Nadeem, Heinrich Peuser, Juliane Richter, Daniel Schäfer and Martin Weigel.

This thesis would also not have been possible without the support of my fellow research group members and–most importantly–friends: Niloofar Dezfuli, Mohammadreza Khalilbeigi and Roman Lissermann. Working with you is a pleasure and it is people like you who make the saying "Find a job you love and you'll never work a day in your life" most reasonable. Thank you.

Last but not least, I am deeply grateful for the love and support of my dearest friends Attila, Martin, Ninon, Paul and my family Kurt, Inge and Katharina.

Contents

CHAPTER 1

Introduction

1.1 Motivation

In the last few years, the capabilities of mobile devices have increased considerably. The mobile phone as-is has undergone an evolution: from originally being a communication device, toward being both, a communication and a sophisticated *information* device (see Jones and Marsden [2006], pp. 9-10). Devices such as the Apple iPhone allow users to access a variety of multimedia data while being on the move. For instance as of writing this thesis, YouTube claims that users upload more than 72 hours of video every minute and over 3 billion hours of video are watched each month (cf. [YouTube, 2012]). Inline with this trend is also the significant development of iTunes U [2012]: a service by Apple, which allows teachers to distribute video lectures and other course materials to students across the globe. With more than 1000 universities participating and about 300 million downloaded video lectures in 2011 (cf. Dalrymple [2011]), it paves the way for empowering students to learn virtually anywhere and anytime; let alone that one third of the access was from mobile devices.

At the same time, these devices generate a need for efficient information exploration and rendering support. While the latter is being addressed with recent advancements in high definition displays, interfaces for mobile multimedia information exploration are in their infancies. Mobile devices impose fundamentally different challenges on the interface design than e.g. traditional desktop computers. This is due to their limited form factors, as well as being used in diverse settings (e.g. being on the move). Furthermore, mobile devices implement novel, emerging interaction paradigms. Both *imposed challenges* and *novel interaction paradigms* render the transfer of existing user interface concepts as-is complicated, if not impossible, and it is thus unclear how to design for the mobile interaction with large multimedia information spaces.

1.1.1 Challenges for Designing Mobile Interaction

Using mobile devices for accessing larger multimedia information spaces imposes various challenges on both designers and users, since mobile devices have severe restrictions due to their form factors (e.g. small displays). This dramatically decreases the amount of information, which can be displayed visibly at once. In turn, this increases the amount information located *off-screen*, since it cannot fit within the small display canvas.

A number of remedies to overcome limited display space were already proposed, yet showed only limited success as two examples shall illustrate: when for instance browsing larger information spaces with a zoomable user interface [Beard and Walker, 1990, Hornbaek, Bederson, and Plaisant, 2002], users need to pan and zoom quite a lot to navigate the space. This can easily lead to losing one's orientation at a certain point (this phenomenon is called *desert fog* [Jul and Furnas, 1998]). Another example is the navigation of hierarchical menus on mobile devices: due to space limitations, typically only one hierarchy level is being displayed at a time. Thus navigating deeper hierarchies with several levels can also lead to the loss of orientation, or even worse, to the loss of the internal locus of control.

Another strikingly challenging aspect of mobile interaction is the usage scenario: mobile devices are used in mobile settings; either in a nomadic fashion, roaming from point *A* to *B*, or in a truly mobile environment, i.e. when being on the move. In these settings, users have to deal with a lot of noise, limited interaction space and high dynamics demanding their attention. As Jones and Marsden put it: "Even when a device is easy to operate and plainly communicates the effect of interacting with it, the test comes when it is deployed in the complex, messy world of real situations [...] when it has to be used in the wild, as it were, in tandem with the world around it, the usability can break down quickly" (see [Jones and Marsden, 2006], p. 51).

1.1.2 Novel Interaction Paradigms

In order to address the specific needs of interfaces for mobile devices, [Gong and Tarasewich, 2004] have developed a set of mobile interface guidelines. These mainly focus on classical graphical user interfaces and follow the well-established windows, icon, menu and pointer (WIMP) paradigm. However, today's mobile devices offer support for a variety of novel forms of input, which go beyond classical graphical user interfaces and each on their own define new interaction paradigms. This comprises for instance touch-based input, which is vastly used on today's mobile phones and allows for touch gestures to control user interface widgets. Another example is the physical input using built-in sensors such as accelerometers (e.g. for tilt-based interaction).

Built-in sensors in combination with optical tracking approaches can also be used to enable a mapping between virtual and physical spaces, as described by Fitzmaurice [1993] in his seminal work on information lenses. By using mobile devices as see-through devices, virtual information spaces are overlain over the physical space. Users can explore the virtual artifacts through embodied interaction: e.g. by pointing the mobile phone into the physical space.

A more recent trend has been picked up by the introduction of mobile hand-held projectors, also called pico projectors [Dachselt et al., 2012]. Combined with or even built into e.g. a mobile phone, they can help in enlarging the display space toward the physical space.

In summary, mobile devices challenge the traditional way we interact with digital information. Existing best practices for classical graphical user interfaces cannot be transferred as-is. This thesis aims at exploring how the affordances of mobile devices can be leveraged from both a user and a designer perspective to design novel visualizations and interaction concepts for large multimedia information spaces. It particularly considers (1) imposed challenges due to form factors (small displays in particular) and input restrictions, (2) the usage in noisy real-world settings and (3) novel interaction paradigms, emerging as an urgent need of (1) and (2). More precisely, it is not the objective of this thesis to serve as yet another set of design recommendations. This thesis investigates particularly promising subsets of interaction paradigms in the scope of large multimedia information spaces, approaches them from an empirical perspective and shows how ecological insights gained in field studies can be harnessed to design interfaces and interaction concepts that are more usable and enjoyable than the state of the art.

1.2 Contributions and Thesis Structure

As argued above, the overarching goal of this thesis is to explore novel, more usable and enjoyable ways for interacting with large multimedia information spaces on mobile devices. Key to this is putting an emphasis on empirical field work throughout the thesis: both in theory generation, as well as in theory verification and evaluation with users. In terms of disciplines, this thesis is situated at the intersection of Human-Computer Interaction, Interaction Design and (mobile) Multimedia. It follows three main research directions (cf. Figure 1.1) which are inspired by the recent technological advancements as outlined in Section 1.1.2.

The first research direction, *device-centric interaction*, looks at mobile devices in an isolated way and concentrates on designing interfaces which can be used efficiently in spite of the small screen, leveraging input modalities such as touch or tilt.

The second research direction, *space-centric interaction*, pushes the boundaries of the virtual information space beyond the small screen toward the physical space. It particularly looks at how mobile devices can be used as see-through displays when mapping the virtual information space to the physical space and how this can improve the usability in mobile settings.

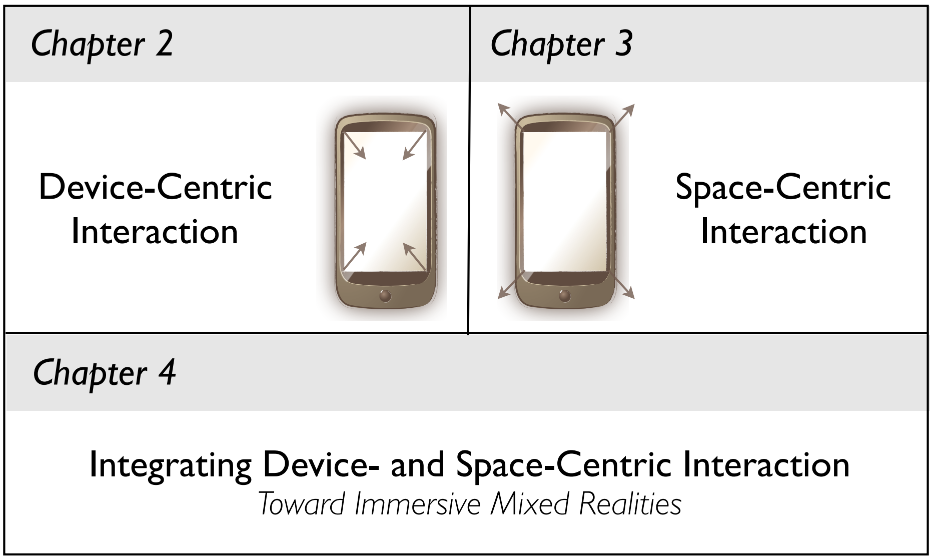


Figure 1.1: Research directions and thesis structure.

In the third research direction, this thesis investigates how virtual and physical space can be tightly integrated, allowing for both device- and space-centric interaction with real world objects–not just one particular mobile device. For this purpose, pico projectors are used to project virtual information spaces into physical spaces, turning objects and planes in arbitrary environments into dedicated projection surfaces and tangible interaction devices.

The three research directions correspond to the structure of this thesis: each direction is discussed within one chapter. Chapter 2 focuses on device-centric interaction, chapter 3 on space-centric interaction and chapter 4 on integrating both interaction perspectives, respectively. Each chapter contains a critical analysis and discussion of related work in the corresponding field of research. Chapter 5 summarizes the outcomes of this thesis and provides an outlook upon future research directions. The contributions within the three main research directions are outlined in the following.

1.2.1 Device-Centric Interaction

In this research direction, we investigate how user interfaces for mobile devices–with their inherent restrictions–can be designed in such a way that they are more usable and provide a richer user experience (cf. Figure 1.2). The research agenda is based upon a guiding scenario for the mobile interaction with large multimedia information spaces: the mobile browsing of topically inter-related video collections.

Design Space Exploration. To advance interaction with mobile video browsers, we have modeled the design space for mobile video browsing. For this purpose, we have *empirically* explored potential user interface designs by means of *two participatory design sessions*. The exploration of these interface types allowed us to set up the *design space of mobile video browsers*. This fills a void in previous research, which was rather fragmentary, and contributes a *fundamental understanding of the design space from a broader perspective*.

Novel Interface Concepts. The exploration of the design space enabled us to *systematically derive 8 interaction concepts* (7 novel concepts, one standard interface), which are situated within the design space and implemented on the

Figure 1.2: Device-centric Interaction. The interface focuses on the inherent device restrictions such as the small screen. The photo shows the *2D Flick* interface, one of the novel interface concepts described in Section 2.5.2.

iPhone. This contrasts GUI-inspired interaction concepts with more innovative concepts for mobile devices, such as gesture-based and physical interaction. They cover the navigation within individual videos, larger videos and for browsing collections of several inter-related videos.

In-depth Evaluation. We carried out an *in-depth evaluation* of these interface concepts, allowing for a broad comparison and a deep understanding of their respective advantages and pitfalls. We conducted a controlled experiment with 44 participants and collected and analyzed more than 18 hours of video observations. Therefore, we were not only able to *assess the usability and user experience* of each interface, but also to identify where errors occur. The results provide *empirical* evidence that designers should leverage the novel capabilities of mobile devices, such as direct touch and inertial sensors.

Detailed Usability Error Analysis. Based on an *ascriptive analysis* of more than *18 hours of video observations*, we evaluated usability criteria and performed a systematic analysis of usability errors. This analysis is both *more comprehensive and more detailed than prior work*. To state only one example, our analysis shows that the video player included in Apple's iPhone has significant drawbacks.

Design Principles. Based on our findings from the in-depth usability evaluation and the detailed usability error analysis, we contribute *orthogonal principles* for the design of future mobile video browsers. This allows designers to build upon our experience and design interfaces that are both *more usable and more enjoyable*.

1.2.2 Space-Centric Interaction

This research direction pushes the boundaries of the virtual information space toward the physical space. For this purpose, we regard the virtual information space as an overlay over the physical space. Thus, the large virtual information space mapped to the physical space. Digital information can be revealed by utilizing mobile phones as see-through devices (for so-called *peephole interaction*, cf. Figure 1.3). The built-in sensors (visual and intertial) recognize the scene and display the corresponding piece of digital information on the mobile phone's screen. Hence, only a small portion of the large virtual information space is visible at a time. Users can explore the virtual artifacts through embodied interaction: e.g. by pointing and moving the mobile phone in physical space.

Empirically Grounded Theory for Embodied Peephole Interaction. The analysis of prior work in this field showed that most of the approaches were evaluated in lab settings and not in a real world context, for which they are actually intended. Thus, the field lacked a fundamental understanding of *how users would actually interact through embodied peepholes in mobile, real world settings*. To address this shortcoming, we conducted a *qualitative, exploratory eld study*. We therefore derived and contribute an *empirically grounded theory*, characterizing embodied peephole interaction in four inter-related categories.

Design Implications. The theory allows us to contribute *design implications* for future embodied peephole interfaces. We demonstrate the application of these implications through the exemplary implementation of *novel interaction techniques* in the field of mobile embodied peephole interaction. As a result of our study, one apparent issue for embodied peephole interaction was the targeting of digital artifacts located off-screen.

Novel Movement Model for Embodied Peephole Navigation. To address the aforementioned issue of targeting digital artifacts located off-screen, we contribute an *empirically grounded mathematical model for the embodied peephole navigation in one-dimensional information spaces*. It models the navigation time to a target in one-dimensional information spaces, depending on the size of the display and the distance to the center of the target. The model is *inspired by physiological aspects of the human body* and thus particularly addresses the affordances of embodied interaction with such displays.

Figure 1.3: Space-Centric Interaction. The virtual information space is laid out in physical space. The user utilizes the mobile phone as a see-through device and points into physical space to reveal the virtual information space.

Empirical Model Validation. We conducted an extensive controlled experiment with 32 participants to empirically validate our mathematical model. We contribute the results and contrast our model with existing ones. We show that it satisfactorily models the movement time in a-priori known spaces–as good as existing approaches. However, in the case of an unfamiliar information space, the results provide significant evidence that the search time is better modeled by our novel formulae.

1.2.3 Integrating Device- and Space-Centric Interaction

The third research direction investigates how virtual and physical space can be tightly integrated, hence allowing for both device- and space-centric interaction with real world objects–not just one particular mobile device. For this purpose, we use pico projectors as so-called "light beams": everyday objects sojourning in a beam are recognized by the system and turned into dedicated projection surfaces and tangible interaction devices (cf. Figure 1.4). This way, our daily surroundings get populated with interactive objects, each one temporarily chartered with a dedicated sub-issue of pervasive interaction.

Conceptual Framework. We contribute the *conceptual framework of LightBeam* and relate it to prior research on pico projectors. Our analysis shows that the tangible character of real world objects has not yet been systematically explored for pico projector interaction. Moreover, it is unclear how the mobility of physical objects could be actually leveraged for tangible interaction.

Qualitative, Exploratory Field Study. In order to develop a fundamental understanding for the problem space, we conducted a *qualitative, exploratory eld study with 8 interaction design researchers*. This allowed us to systematically explore the light beam concept. We contribute our findings, which identify relevant theoretical dimensions comprising the projector placement and how to provide both in- and output with real world objects.

Novel Interaction Techniques. We advance the field of pico projector interaction by designing *novel interaction techniques* which go beyond prior work. To state one example, our study results indicated that the problem of overloading mappings of physical objects can be solved by spontaneous overloading. For this purpose, we map the unique affordances of everyday objects such as rotating to digital functions. We provide a loose coupling of interaction and object, since any object which affords the appropriate manipulation (such as rotation) can be used to carry out that very function. The mapping is not directly bound to one particular physical object.

Evaluation. The implemented interaction techniques were evaluated in an *early user feedback session*. We contribute the results, which confirm the identified theoretical dimensions from the first study and underline the importance of the tight integration of both virtual and physical space.

Figure 1.4: Integrating Device- and Space-Centric Interaction. LightBeam, the camera-projector prototype contributed in Chapter 4, is shown at the bottom of the picture. Objects sojourning in the beam, here a card box, are turned into dedicated projection surfaces and tangible interaction devices.

1.3 Research Methodology

The three research directions each follow essentially the same research methodology (see Figure 1.5 for a schematic overview). Based upon a thorough analysis of existing work in the research field, the innovative part of the methodology can be subdivided into two main parts: (1) an inter-twined qualitative exploration and theory generation process and (2) a cyclic design process. Each part is described in the following.

Qualitative Exploration and Theory Generation. Mobile interaction takes place in the field, a "complex and truly messy environment" (cf. section 1.1.1). The interaction itself therefore highly depends on its context of use. To better understand how technology such as mobile devices would be used in these contexts, we suggest to carry out qualitative explorations in the field with potential users of the technology. Our agenda is therefore inspired by the *situated action approach* by Suchman [1987] and other modern theoretical approaches, such as *grounded theory* [Glaser and Strauss, 1967, Strauss and Corbin, 2008, Sharp et al., 2007]. The qualitative exploration uses ethnographic methods, such as

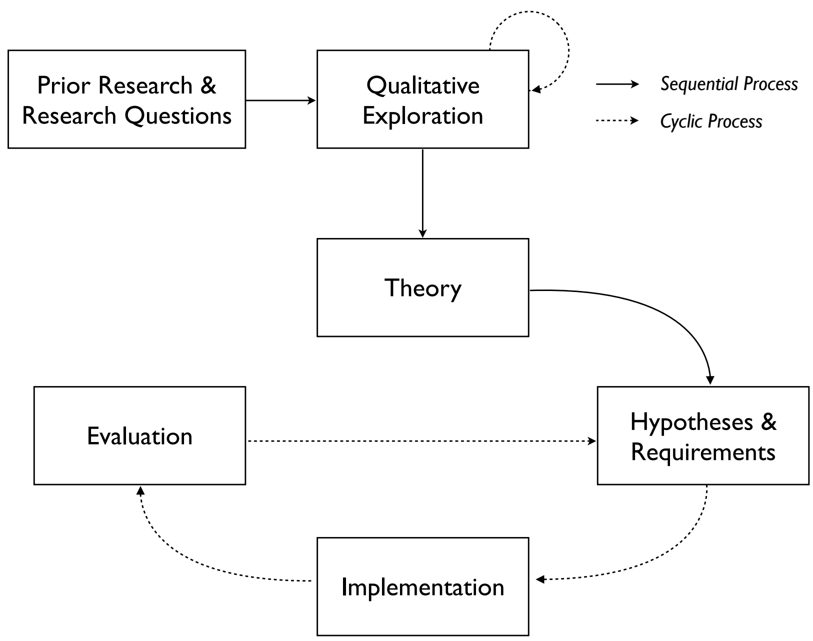


Figure 1.5: Overview over the research methodology.

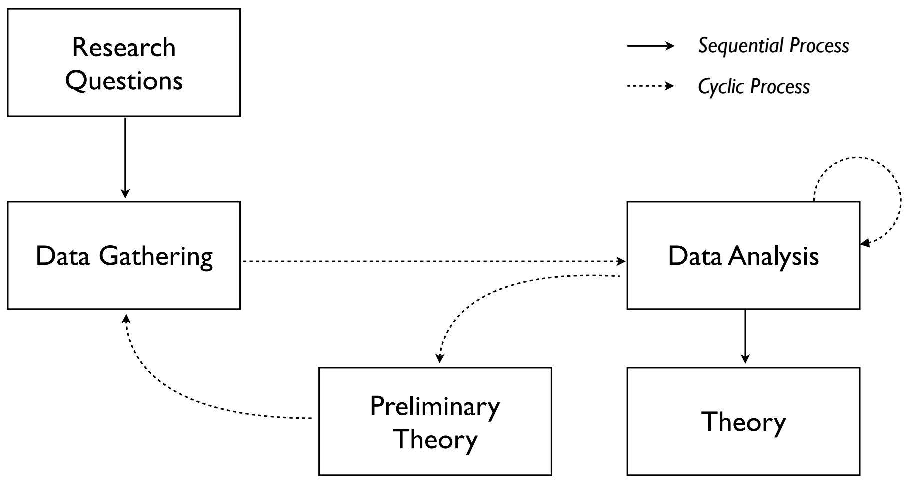


Figure 1.6: Schematic overview over the iterative theory generation process.

observation, interviews, photo documentation or note taking to identify relevant *phenomena in real world use contexts*.

The main objective of the qualitative exploration is theory generation. We hereby refer to theory in the sense of Rogers [2012] (p. 4): "a *theory* is a well-substantiated explanation of some aspect of a phenomenon". Our methodology foresees to employ an *iterative theory generation process* (see Figure 1.6 for a schematic overview), consisting of an iterative data gathering and data analysis methodology. This allowed us to refine the theory and deeply explore particularly relevant phenomena. Examples of this are the participatory design process outlined in section 2.3, where different lo-fi prototypes were used in different situations to broadly explore the design space, and the exploratory field studies in section 3.2 and section 4.3.

Cyclic Design Process. In terms of chapter 2 and 4, our methodology suggests a user-centered design approach [Norman and Draper, 1986]. Central to user-centered design is the involvement of potential users in both the design and the analysis. As an initial input into the process, the methodology foresees to use the generated theory, explaining phenomena observed in the field. This informs the initial design in terms of hypotheses and requirements.

After implementing the initial design of the corresponding user interfaces, we suggest to analyze the prototypes rigorously in a variety of studies. In terms of chapter 2, we conducted an extensive controlled experiment to compare and contrast the usability and user experience of the implemented interfaces. In chapter 4, we conducted early user feedback sessions to get an initial impression on the usability, as well as the user experience of the designed interactions.

In terms of chapter 3, we identified an apparent phenomenon during the qualitative exploration, the problem of targeting digital off-screen artifacts in embodied peephole pointing. This then lead to the development of a theoretical model and the derivation of various hypothesis. This allowed us to implement an actual physical prototype, which in turn enabled us to conduct a controlled experiment to validate the model and to test the identified hypotheses.

In all cases, the results of the analysis process then directly inform the requirements for future interface designs, closing the cycle of the design process.

1.4 Publications

Parts of this thesis are published in proceedings of international conferences such as ACM *Multimedia*, ACM SIGCHI Conference on Human Factors in Computing Systems (*CHI*), BCS Conference on Human-Computer Interaction (*HCI*), IEEE International Conference on Advanced Learning Technologies (*ICALT*) and ACM SIGCHI Conference on Mobile and Ubiquitous Multimedia (*MUM*), as well as international workshops.

Parts of the chapter on device-centric interaction are published in [Huber et al., 2010a,b,c,d,e]. The exploratory field study on dynamic peephole pointing is published in [Huber, 2010]. The model of embodied dynamic peephole pointing is published in [Huber et al., 2011b]. Parts of the chapter on bridging both device- and space-centric interaction are published in [Huber et al., 2011a, 2012a,b].

The mobile device sketches in Figure 1.1 are re-printed with permission[1].

[1] http://www.eleqtriq.com/ (last checked: October 29, 2012).

CHAPTER 2

Device-Centric Interaction

In this chapter, we will approach mobile interaction with large multimedia information spaces from a *device-centric* interaction perspective. We will investigate how user interfaces for mobile devices, being restricted to both a small screen and limited input capabilities, can be designed in such a way that they are more usable and provide a richer user experience. This chapter focuses on mobile browsing of topically inter-related video collections as a guiding scenario for the mobile interaction with large multimedia information spaces. We will therefore explore the design space of mobile video browsers, present various novel interfaces and report on their usability, as well as their user experience. Based upon these results, we will derive design implications for future interfaces.

This chapter is structured as follows. In Section 2.1, we outline challenges for designing user interfaces for mobile video browsing with respect to use patterns, such as learning or knowledge work. The analysis shows that it is important to support different navigation complexities, such as the navigation of inter-related video collections, and not only the interaction with individual videos or even only small video segments. Section 2.2 reviews related work. In order to advance interaction with mobile video browsers, we have modeled the design space for mobile video browsing. For this purpose, we have initially explored potential user interface designs with respect to the identified navigation complexities by means of two participatory design sessions, which are described in Section 2.3. The exploration of these interface types allowed us to set up the design space of mobile video browsers in Section 2.4. The design space is spanned alongside two dimensions: *navigation complexity* and *interface type*. The latter covers GUI-inspired interaction concepts and more innovative concepts for mobile devices, such as gesture-based and physical interaction. Section 2.5 then illustrates the 8 final interface concepts (7 novel concepts, one standard interface), which are situated within the design space and implemented on the iPhone. An in-depth evaluation of these interfaces is presented in Section 2.6. This allows for a broad comparison and a deep understanding of their respective advantages and pitfalls. Based on an ascriptive analysis of more than 18 hours of video observations, we evaluated usability criteria and performed a systematic analysis of usability errors. This analysis is both more comprehensive and more detailed than prior work. To state only one example, our analysis shows that the video player included in Apple's iPhone has significant drawbacks; it also points out how future versions could be designed to be more efficient to use. Based on our findings, we present principles for the design of future mobile video browsers in Section 2.7. The chapter concludes with a discussion of the results and an outlook upon future work in Section 2.8.

In summary, the main contributions of this chapter are

- exploration of the design space for mobile video browsing,
- novel interface concepts,
- an in-depth evaluation of these concepts with a detailed usability error analysis and
- design principles for future mobile video browsers.

2.1 Use Patterns and Challenges of Mobile Video Browsing

Increasingly powerful mobile devices like the Apple iPhone are currently dramatically changing how we perceive multimedia when being on the move. Users are able to access a constantly increasing number of video streams almost anytime and anywhere. Videos are not only watched for entertainment during leisure time [O'Hara, Mitchell, and Vorbau, 2007], but also used at work, e.g. for learning on the job or for mobile learning [Hürst, Welte, and Jung, 2007b].

In the following, we illustrate that requirements for mobile video browsing depend on the complexity of the browsing task. In light of different use patterns, we show that these aspects require

- getting detailed information on the current topic contained within an *individual video segment*,
- an efficient overview over a *large video* with quick and easy access to any of the contents and
- quick and easy navigation to information which is related to the current topic within a *collection of inter-related videos*.

2.1.1 Detailed Information in Individual Video Segments

Using mobile devices for watching videos reveals various challenges: on the one hand, mobile devices have severe restrictions due to their form factor (e.g. small displays), but on the other hand they also offer support for novel forms of input (e.g. direct touch and physical input). Moreover, users in mobile settings typically cannot devote their full attention to the user interface. Hence, high efficiency and effectiveness of the user interface is crucial [Gong and Tarasewich, 2004]. Particularly when watching individual video segments (i.e. short videos of e.g. 5 minutes in length) on mobile devices, phenomena like *overshooting* can occur. In the case of overshooting, a user skims the video for a particular scene, but misses the scene she was actually looking for, e.g. because the playback speed was too fast. This then prevents users from getting the desired information within that very video segment, prolonging their navigation task and therefore resulting in a decreased usability.

Some approaches for novel and efficient user interfaces for mobile video browsing have been presented [Karrer, Wittenhagen, and Borchers, 2009, Sun and Hurst, 2008]. While these are valuable contributions for specific aspects of mobile video use, research is still very fragmentary, pointing on individual aspects of the design space. The field still lacks a general understanding of the design space for mobile video browsing and of the characteristics of specific interaction concepts. A precise knowledge of the advantages and pitfalls of different interface concepts, of frequent use patterns and of recurrent misconceptions related to these concepts is likely to significantly increase the quality of future mobile video browsers.

In order to set-up the design space of mobile video browsing, it is important to understand the patterns of use. O'Hara et al. [2007] studied how mobile devices are used for consuming videos for the purpose of entertainment. The study shows that patterns of mobile video consumption differ from just "watching TV" in a mobile settings. The authors present a range of deeper motivations related to the social setting. Mobile video browsers are used for sharing video contents by watching them with other people in social situations. In contrast, they are also used for being in proximity to others (e.g. with family at home) while not disturbing them with own content. Moreover, mobile video can also be characterized as a "privatizing technology": by watching video contents, users can withdraw themselves from public space when being in the proximity of other people, e.g. in public transport.

2.1.2 Efficient Overview Over a Large Video

Mobile videos browsers are not only used for entertainment, but also for mobile learning and training on the job. The ubiquitous availability of multimedia learning material through services like iTunes U [1] or OpenCourseWare [2] has paved the way for groundbreaking changes in mobile learning. A recent study [Hürst et al., 2007b] found a shift in the usage habits of students towards using the mobile version of lecture recordings, e.g. when commuting [Schwanen, 2008]. These settings of learning and working result in different use patterns than entertainment.

Much more than entertaining videos, instructive videos are watched in a non-linear manner. Analogously to the work with textbooks, users tend to watch specific passages of interest and jumping between different passages instead of linearly consuming the entire video, particularly in case of large videos, e.g. lengthy lecture recordings.

2.1.3 Easy Navigation In a Collection of Inter-related Videos

Besides watching individual videos, the interrelationship of several videos (e.g. as hyperlinks in so-called hypervideos) is of major importance for successful learning processes and trainings. The relationships are crucial for contrasting and integrating knowledge which is contained in related videos. This can

[1] http://www.apple.com/education/mobile-learning/ (last checked: October 29, 2012).
[2] http://ocw.mit.edu (last checked: October 29, 2012).

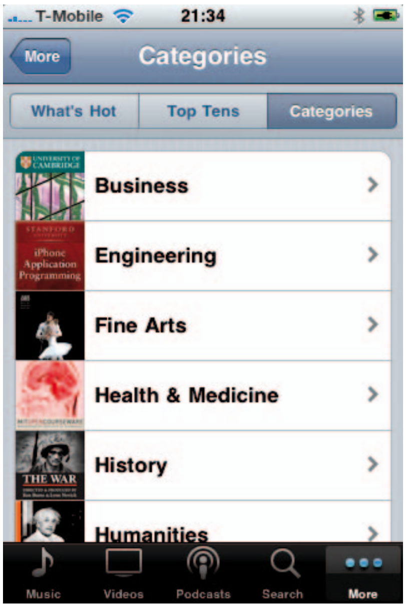


Figure 2.1: iTunes U digital library browser for the iPhone OS. Users can either search for lectures or choose from various categories. However, lectures can only be watched as an ordinary movie.

be compared to reading books and articles, where we follow references and compare and integrate information from various documents. For example, various topically related videos from different institutions allow learners to receive elaborate explanations for a certain problem and can be used to gain deeper insight into a specific problem domain from a slightly different point of view. This practice is possible nowadays due to the vast amount of educational videos available online from various universities. However, state of the art mobile video browsers do not support the user sufficiently in these tasks, which involve the use of multiple lecture recordings. A learner would have to (1) identify potential lectures in the digital library browser (see Figure 2.1), (2) scan each lecture sequentially to check whether it really covers the desired topic and (3) note down or memorize the occurrences and correct positions within the e-lecture. Hence, it is impossible for learners to complete this task in a reasonable amount of time in a mobile setting.

In summary, patterns of mobile video browsing include using several inter-related videos. Hence, not only the efficient navigation within an individual video, but also within collections of several inter-related videos is crucial for video browsing, e.g. in knowledge work.

2.2 Related Work

In the following, we discuss related work in the field of mobile video browsing. While the core of this section focuses on mobile interaction, we first revisit desktop-based interfaces. There exists a large body of knowledge on how to design interfaces for video navigation on desktop computers. However, deriving implications for mobile interfaces based on desktop-based variants and direct adaption of lessons learned in this very design space are hardly possible. The input modalities, as well as the physical constraints (such as screen real estate) are fundamentally different. But at the very least, desktop-based interfaces serve as a source of inspiration to many mobile interfaces.

2.2.1 Desktop Interfaces

Desktop interfaces are typically GUI-based, rely on traditional input modalities like a keyboard and a mouse and are designed with the thought of having a large display real estate at hand. While there is a plethora of different approaches to designing desktop-based interfaces, we shortly touch upon the most relevant interaction techniques for our agenda: (1) direct manipulation for fine-grained in-scene navigation, (2) non-linear video segment navigation and (3) hypervideo and hierarchical video browsing for video collections.

2.2.1.1 In-scene Navigation through Direct Manipulation

Fine-grained in-scene navigation is often performed by dragging a slider along the timeline–a widely used user interface element in today's media players such as iTunes or the Windows Media Player. Dragging it to the right advances and to the left goes back in time respectively. A more sophisticated approach explored by Dragicevic et al. [2008] and also by Karrer et al. [2008] employs direct manipulation. Here, the user directly interacts with objects in a video, allowing for a frame-accurate navigation. In the case of Figure 2.2, the user can navigate within that very video scene by dragging the billiard ball along its trajectory. Dragging it forth and back then translates to advancing and going back in time. This technique is particularly useful when performing a very fine-grained navigation task; when for instance analyzing a video frame by frame.

Since objects are mostly present in a particular scene, applying this technique to a more coarse-grained navigation within larger videos or even across interlinked video collections is impossible as-is.

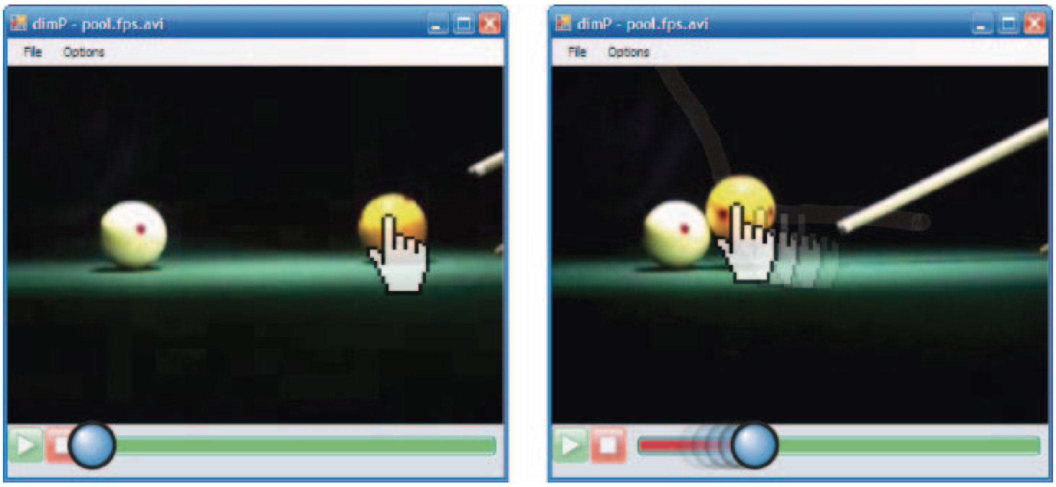


Figure 2.2: The Direct Manipulation Video Player presented by Dragicevic et al. [2008]. Here, the user can navigate within the scene by dragging the billiard ball along its movement trajectory (as seen on the right hand side).

2.2.1.2 Non-linear Video Segment Navigation

There also exists a variety of non-linear browsing techniques, supporting navigation tasks in video segments. A prominent example is the LEAN system by Ramos and Balakrishnan [2003]. Besides video annotation being its primary focus, it also embodies fluid interaction techniques for video navigation. In particular, it contributes the PVSlider, which employs a rubber band metaphor for timeline-based navigation (see also Hürst et al. [2004] for a similar implementation). The timeline control knob is mapped to a rubber band. By dragging the knob, the user spans the rubber band and therefore the playback speed is adjusted adaptively. The stronger the band is spanned, the faster the video playback. The system itself is conceived for stylus-based input. This inspired various mobile interfaces, being also stylus-based, which are discussed in the following subsection.

There have also been efforts in exploring techniques which explicitly support fast-forwarding of video segments. For this purpose, Cheng et al. [2009] adopt a metaphor of scenic car driving for video fast-forwarding: when the user approaches interesting regions of interest within a video segment, the video player shall slow down and on the contrary speed through uninteresting regions of interest. As shown by Höferlin et al. [2011], this can be particularly helpful for browsing video surveillance material. More recently, Schoeffmann et al. [2010] investigated efficient navigation techniques for single videos based on fast content analysis. The interface primarily is centered around the idea of navigation summaries. A summary can be a collection of keyframes describing a region of interest in the video. The interface features one or more timelines with different visualizations, e.g. abstract representations of semantically similar video segments. This supports a fast content-based search, while requiring a certain amount of screen space for different visualizations and timelines.

2.2.1.3 Hypervideo and Hierarchical Video Browsing

There are also quite a few prominent examples enabling hypertext-like interaction with so-called *hypervideos* in inter-related video collections. In this vein, a larger body of research [Hürst and Götz, 2004, Brotherton and Abowd, 2004] and commercial products (e.g. Lecturnity [2012]) are concerned with capturing, browsing, searching and playback of educational videos. Well-known examples are the Cornell Lecture Browser [Mukhopadhyay and Smith, 1999] and the Berkeley Internet Broadcasting System (BIBS) [Rowe et al., 2001]. Both systems enable students to access and play back video lectures in a web browser. In the case of BIBS, lectures can be played back both live and on demand, with slides, slide indices and keyword search functionality being displayed in different windows. BIBS allows students to browse a large collection of video lectures through a hypertext-based program guide.

Similar to BIBS is the Microsoft Research LecCasting System (MSRLCS) [Zhang et al., 2008], a fully automated lecture capture and broadcasting system. MSRLCS supports live broadcasting, on-the-fly slide capture and minimal pre- and post-production times. Part of MSRLCS is iCam2, an automated lecture content capturing system. The collection of lectures can be browsed using a calendar-based interface and individual lectures can be navigated using the timeline of a web-based video browser or by selecting individual slides. Like BIBS, MSRCLS supports both live and on demand playback.

Figure 2.3: Example screenshot of the MIT lecture browser. Students can search for lectures through a keyword-based search at the top. A media player enabling video playback can be found on the right hand side. The transcript of the lecture is located at the bottom right. Related keywords in other lectures or their respective transcripts are visualized in the center. [Glass et al., 2007]

Another prominent example is the MIT spoken lecture processing project [Glass et al., 2005, 2007]. It is a web-based video browser for lecture recordings, enabling students to not only browse and watch lectures online, but it also provides automatic speech transcription. All sources are synchronized and students can navigate in either of them. Hence, clicking onto a word in the transcript in turn also navigates in the lecture recording (cf. Figure 2.3).

More recently, Adcock et al. [2010] presented TalkMiner, a rich search and browsing system for existing lecture webcasts without a-priori instrumentation or post-capture authoring. It scaffolds students in finding and accessing specific topics in video lectures efficiently. Students can then browse the selected video lecture using a slide overview and an embedded video browser.

Other application domains for hypervideo interaction in inter-related video collections are for instance news broadcast archives. A prominent example is the Informedia project [Wactlar et al., 1996] conducted for over 10 years at Carnegie Mellon University [Hauptmann, 2005]. The project mainly focused on search, retrieval, visualization and summarization of video archives.

Concerned with hypervideo interaction in general, HyperHitchcock consists of an authoring and a navigation tool [Shipman et al., 2008]. In the latter, Shipman and colleagues visualized related videos along the video timeline, showing a keyframe of the video as a description. Furthermore, HyperHitchcock provides video summaries (a concatenation of salient regions of interest per video clip) at any level of the video hierarchy.

VastMM by Haubold, Dutta, and Kender [2008] is both a video indexer and a video browser with various features, such as video annotation, bookmarking and multiple visualization. The latter i.a. provides video abstraction through keyframe visualization, connecting thumbnail keyframes, as well as visual annotation to the video timeline. VastMM was mainly used as an instrument in various studies to understand the effect of cues (auditory, visual or textual) on video search tasks. Comparable to HyperHitchcock and the MIT spoken lecture processing project, it requires a certain amount of screen space for the various different visualizations, shown side-by-side.

More recently, Del Fabro, Schoeffmann, and Böszörmenyi [2010] have explored non-sequential hierarchical video browsing. Hierarchical browsing here means to browse different abstraction levels of a video, e.g. starting with a segmented video and browsing toward the keyframe level. Their tool provides different views, namely a parallel view (showing n equally subdivided parts of a video) and a tree view (where browsing paths are modeled as a tree). In both cases, a user has to go sequentially through the hierarchy levels until the desired part of the video has been found. Schoeffmann and Fabro [2011] have built upon this prior work and designed a 3D carousel view of video hierarchies. This overcomes the problem of having to navigate each level sequentially by showing different visualizations of the different levels. While this is primarily a prototypical implementation, its usability and usefulness has yet to be determined.

2.2.2 Mobile Interfaces

The aforementioned desktop interfaces are space-consuming and require larger screens to allow users to browse e.g. hypervideos efficiently. In contrast, mobile devices offer different affordances leading to novel input modalities such as touch or tangible input. Furthermore, they have fundamentally different form factors. Particularly due to the highly limited screen real estate, hypervideo interfaces cannot be transferred to mobile devices as-is. As pointed out earlier, research on mobile video browsers is rather fragmentary and only punctual contributions have been achieved. To the best of our knowledge, prior work has not contributed a detailed analysis of the design space of mobile video browsers. Instead, prior research mostly focuses on a specific aspect of video browsing such as fine-grained in-scene navigation. Most related approaches are discussed in the following.

2.2.2.1 In-scene and Video Segment Navigation

Most relevant are various interface concepts for both pen and touch-based interaction on personal digital assistants (PDAs) by Wolfgang Hürst et al. Among them is the ElasticSlider [Hürst and Meier, 2008], which allows users to skim quickly through continuous video streams. This approach leverages a rubber band metaphor and is largely inspired by the PVSlider in LEAN [Ramos and Balakrishnan, 2003].

Another interface is the so-called MobileZoomSlider [Hürst, Götz, and Welte, 2007a]: based upon its desktop version, the AV-ZoomSlider [Hürst, 2006], it contributes novel concepts for timeline-based video navigation, allowing users to skim a video on different granularity levels. Here, the timeline is visualized at the bottom of the screen. However, it can be accessed from anywhere: clicking for instance on the top of the screen thus provides a proxy to the timeline and hence, users can navigate essentially anywhere on the mobile device's screen. The MobileZoomSlider differentiates between position-based and speed-based navigation. The position-based navigation maps the vertical position of the click to the scale of the navigation speed. Clicking at the top maps to navigating in the finest possible scale and clicking at the bottom, near the timeline, results in the coarsest navigation scale. The space in between is mapped to a linear interpolation of the navigation scale. The speed-based navigation maps the vertical position of the click to the playback speed (top being the slowest and bottom being the fastest speed). Both navigation modes can be accessed at any time. For this purpose, Hürst et al. have subdivided the interface horizontally. Clicking within a small vertical margin on the right hand side enables speed-based navigation. Clicking anywhere else on the screen switches to position-based navigation.

Last, Hürst et al. have designed a ScrollWheel interface [Hürst, Meier, and Götz, 2008] inspired by the iPod's clickwheel. Comparable to the MobileZoom-Slider, it focuses on manipulating the timeline. Here, they map the timeline onto a circle. The radius of the circle determines the navigation scale. The larger the radius, the finer the navigation since longer distances need to be crossed to advance in the video. Additionally, Hürst and Merkle [2008] have also investigated one-handed mobile video browsing. For this purpose, they have designed an interface inspired by DVD menus, showing keyframe thumbnails for each part of a video segment. A summarizing overview over their interface concepts can also be found in [Sun and Hurst, 2008]. However, none of these concepts support selective interaction or the navigation between collections of inter-related videos (e.g. through hyper-video).

Fine-grained in-scene navigation is also supported by PocketDRAGON [Karrer et al., 2009]. Based upon the desktop version [Karrer, Weiss, Lee, and Borchers, 2008], PocketDRAGON supports time-based video navigation through direct manipulation. When a user touches on an object in a video, e.g. a car, its trajectory within the video is visualized. The user then can drag the object along its trajectory and can therefore navigate forth and back.

2.2.2.2 Other Navigation Concepts

MiniMedia surfer is a mobile browser for small video segments developed by Kamvar et al. [2004]. The browser supports keyword queries and users explore query results through key frames. The navigation completely relies on the designated keywords for each video segment. This is a major issue when trying to get an overview on a set of videos without knowing what to look for.

Similar to MiniMedia surfer, MobileTiles [Falchuk, Glasman, and Glasman, 2008] leverages key frame abstraction to support a user in trying to grasp the contents of a movie. Here, a set of keyframes of a movie is provided to a user, allowing her to get an overview over these keyframes. This can be for instance beneficial to support a user in her decision, whether she wants to watch a movie or not. However, MobileTiles itself does not support video playback.

While not directly applied to video browsing, yet a first step toward novel input modalities, Dachselt and Buchholz [2009] presented tilt-based multimedia interactions for mobile devices. The mobile device is used as a remote control for media displayed on a distant screen. Tilting the device then translates to navigating forth and back through e.g. a photo collection.

In summary, research on mobile video browsers is rather scarce, focusing on a specific aspect of video browsing such as fine-grained in-scene navigation. Toward a more holistic view upon the design space for mobile video browsing, we initially explored potential user interface designs with respect to the identified navigation complexities by means of two participatory design sessions. In the following section, we describe the design process in greater detail.

2.3 Participatory Design Process

In section 2.1, we showed that patterns of mobile video browsing involve three different classes of navigation complexity: navigating (1) an *individual video segment*, (2) a *larger video* and (3) a *collection of inter-related videos*. The review of related work in the field of mobile video browsing revealed that particularly the last two have only been scarcely explored. However, in the scope of mobile video browsing, e.g. for mobile learning or knowledge work, they are very important–and challenging in particular at the same time.

We conducted two participatory design sessions with focus groups to identify potential interface and interaction designs for these levels. The main objective was to get a broader view upon the design space and to elicit design requirements. The goal was then to eventually

1. characterize the design space for mobile video browsing with respect to the aforementioned levels of complexity of navigation and
2. narrow the explored design examples down to final interface concepts, which can then be situated within the actual design space.

The utilized methodology, identified interface concepts and further results are discussed in the following.

2.3.1 Setup, Study Design and Methodology

The participatory design process was conducted in two sessions with focus groups. There were eight participants in total (6 male, 2 female), four participants per session. For each session, we chose one media designer, two computer scientists and one pedagogue as participants. We chose different participants for each session. Their age ranged from 25 to 35 years (27 in average). Each session lasted about two hours.

In both sessions, the participants were first given an overview of the study objectives. Moreover, they were introduced to the concepts of mobile video browsing with respect to the three levels of navigation complexity described above. In the first session, the participants were provided with paper prototypes as stimuli. They were then asked to discuss them and were provided with additional paper templates to sketch their design ideas. We then used the results from the first session to further refine our stimuli. These were in turn used as input for the second session, where the prototypes were further improved.

As data gathering methodologies, we used video recording, observation, photo documentary and both paper prototypes and design sketches from the participants. We transcribed the data and then analyzed salient quotes using an open coding approach.

2.3.2 Session 1: Paper Prototypes

In the first session, we used paper prototypes for the three levels of navigation complexity as stimuli. The paper prototypes were of low, as well as higher fidelity. The low fidelity paper prototypes allowed us to explore certain interface concepts, whereas the those of higher fidelity allowed us to get specific feedback to a more concrete interaction technique.

We used printed iPhone templates in combination with printed prototypes as shown in Figure 2.4. The printed prototypes could then be laid on top of the iPhone template, allowing the participants to imagine how the actual interface would look like. We also had a real iPhone at hand, which the participants

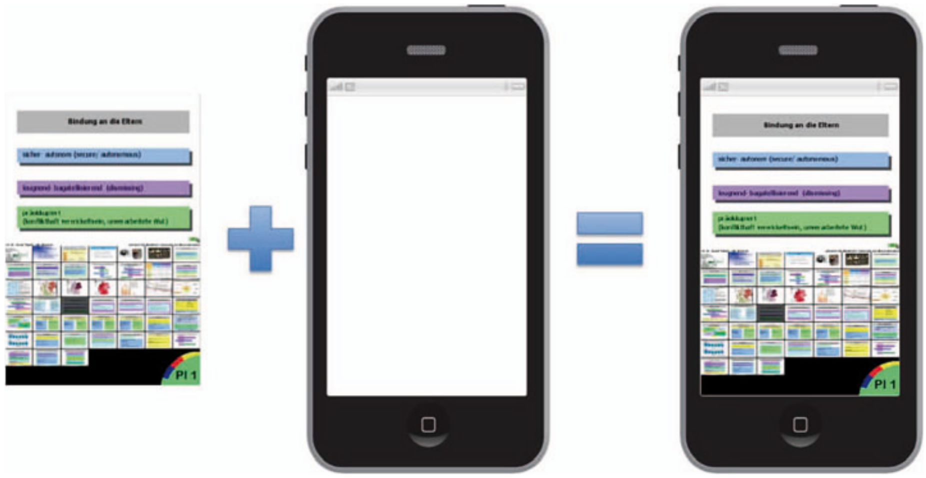


Figure 2.4: Example of a printed paper prototype, which was laid on top of an iPhone paper template.

could also use in combination with the paper prototypes (again, laid on top), to get a better feeling for both device dimensions, as well as weight. We opted to use these templates, since they resemble the iPhone form factor and the actual prototypes can be easily exchanged on top of them.

The paper prototypes were presented and discussed one after another. Once a discussion was finished, the participants were asked to exchange the paper prototype on top of their iPhone template. In the following, we illustrate the presented stimuli.

2.3.2.1 Individual Segment Navigation

The first interface concept targeting the navigation in individual video segments was designed to provide the core functionality of a video player (i.e. playback, timeline-based navigation), while supporting users in mobile situations. Figure 2.5 shows the interface mockup. While a video is being played back, simply touching the screen anywhere pauses the playback (see Fig. 2.5 left). A further single touch resumes playback respectively (see Fig. 2.5 right). The timeline is visualized as a quarter circle in the bottom right corner. Users can directly touch the timeline knob and navigate within the video segment by dragging the knob alongside the circular timeline. The timeline was designed in this very way due to a simple reason: when a video is played back in landscape mode, the timeline should be easily accessible by a user, even during one-handed use.

Figure 2.5: From left to right: the interface concept for the navigation in individual segments, allowing users to directly touch the display for pausing and respectively replaying the video. Moreover, the timeline on the bottom right can be used to navigate through the video.

Figure 2.6: Tilt-based video navigation concept.

In addition to the timeline-based navigation, we have also designed a sensor-based navigation concept. By leveraging the built-in accelerometer of today's mobile devices, a user can navigate within a video (cf. Fig. 2.6): the navigation mode can be enabled by touching the screen, then dragging and holding the appearing user interface element to the right hand side (see Fig. 2.6 top). Once the mode is enabled, the device can be tilted to the right to fast forward and to the left to fast rewind, respectively. The browsing speed is visualized using arrows (see Fig. 2.6 bottom). Here, a continuous manipulation of the navigation speed is assumed, meaning: the more the device is tilted, the faster the navigation speed. The mode is deactivated by releasing the finger from the screen. In addition to the paper prototype, we also presented an early prototype to the participants, which was implemented on the iPhone. This way, the participants could have a first impression on the actual tilt behavior.

2.3.2.2 Large Video Navigation

We have designed two interface concepts for the navigation in larger videos. The major goal here was to support a user's overview over the video and to enable the user to browse larger videos efficiently. Both interface concepts utilize two core concepts: keyframe abstraction (e.g. through temporal segmentation or shot detection), as well as color coding of different videos. Keyframe abstraction is used to support the non-linear navigation of larger videos. The color coding of different videos is used to assign a color to a video, therefore providing shortcut functionality by selecting colors, instead of a particular video e.g. from a list of videos. Concrete realizations of these two concepts are presented below.

The first interface is shown in Figure 2.7. Here, four keyframes are visualized at once, from top to bottom and from left to right. Tapping onto one of the keyframes starts playback. By flicking horizontally over the keyframes, a user can skip to the previous or next four keyframes in the video respectively. The color circle at the bottom right corner designates other videos. The colors serve as a *visual cue of low delity* for the user, to help her remember the existing videos. The concept foresees that a user assigns a color to a video, as soon as she starts browsing it for the first time. By tapping onto the circle and then selecting a color, users can switch to a different video. As soon as the user hov-

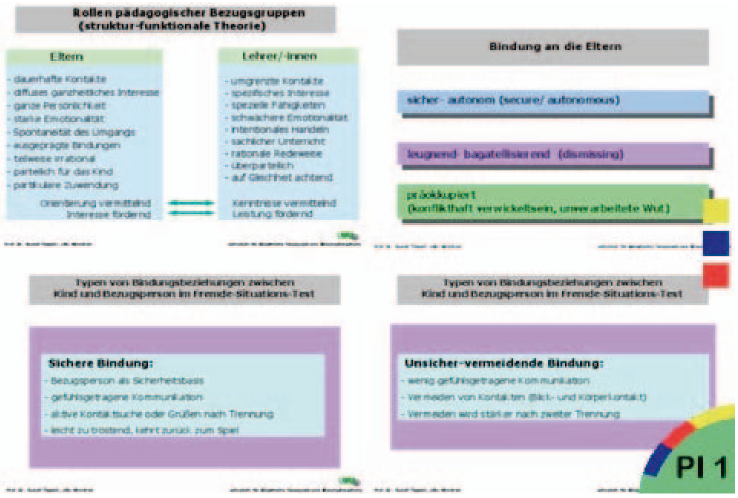


Figure 2.7: Two-dimensional navigation concept.

ers over the colors, video names are visualized. Moreover, colored blocks are visualized on the right hand side of the interface. These designate the browsing history. Here, the user first watched the video associated with yellow, then blue and eventually red. By tapping onto e.g. the blue block, the interface jumps back to the point where the user left the video associated with blue.

The second interface is illustrated in Figure 2.8. The interface is subdivided into two parts and uses an overview+detail metaphor: the upper part shows the current keyframe and the lower part shows an overview over all keyframes. Users can switch between keyframes by flicking horizontally over the upper part. In addition, users can also select keyframes directly in the grid-based layout at the bottom. Since the thumbnails are rather small, users can also zoom in (cf. Fig. 2.8 right) by double tapping onto the grid. One zoom level is supported. Again, the color circle for video selection is displayed in the bottom right corner of the interface.

2.3.2.3 Inter-related Video Collection Navigation

The previous concepts focused on the navigation in either small video segments or larger videos without any semantic inter-relationship. To address the navigation within topically inter-related video collections, we have designed the interface presented in Figure 2.9. In essence, we have enhanced the grid-based interface for the navigation in large videos (as shown in Fig. 2.8) and added

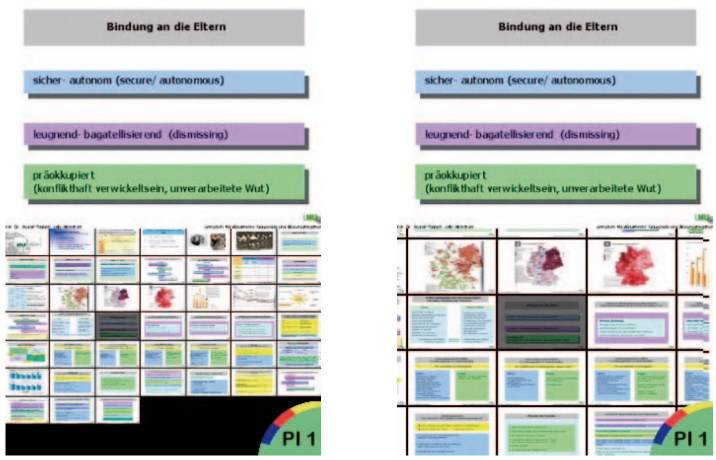


Figure 2.8: From left to right: Grid-based navigation concept (left) and zoomed in grid-based layout (right).

an arrow to the currently displayed keyframe (cf. Figure 2.9 left), whenever related content (i.e. parts of other videos) is available.

By simply dragging the arrow downwards, the interface scrolls down and reveals related content (cf. Fig. 2.9 middle). The related videos are visualized using thumbnails of the videos and the video's title. Moreover, the lines connecting the preview thumbnails to the current keyframe are colored according to the color coding for the videos, which is also available through the color circle at the bottom right. The color circle concept is identical to the one shown in Figure 2.7.

When a user taps onto a preview thumbnail, the interface is scrolled down further and the related keyframes are displayed (cf. Fig. 2.9 right). The colored line is forked into further lines, designating the related keyframes. The currently displayed related keyframe is indicated by the arrow and other keyframes can be accessed by tapping onto the respective lines.

The interface also supports deeper links: in case the related keyframes contained further references to other related videos, an arrow as in Figure 2.9 (left) would be displayed and the related videos could be browsed in the very same manner.

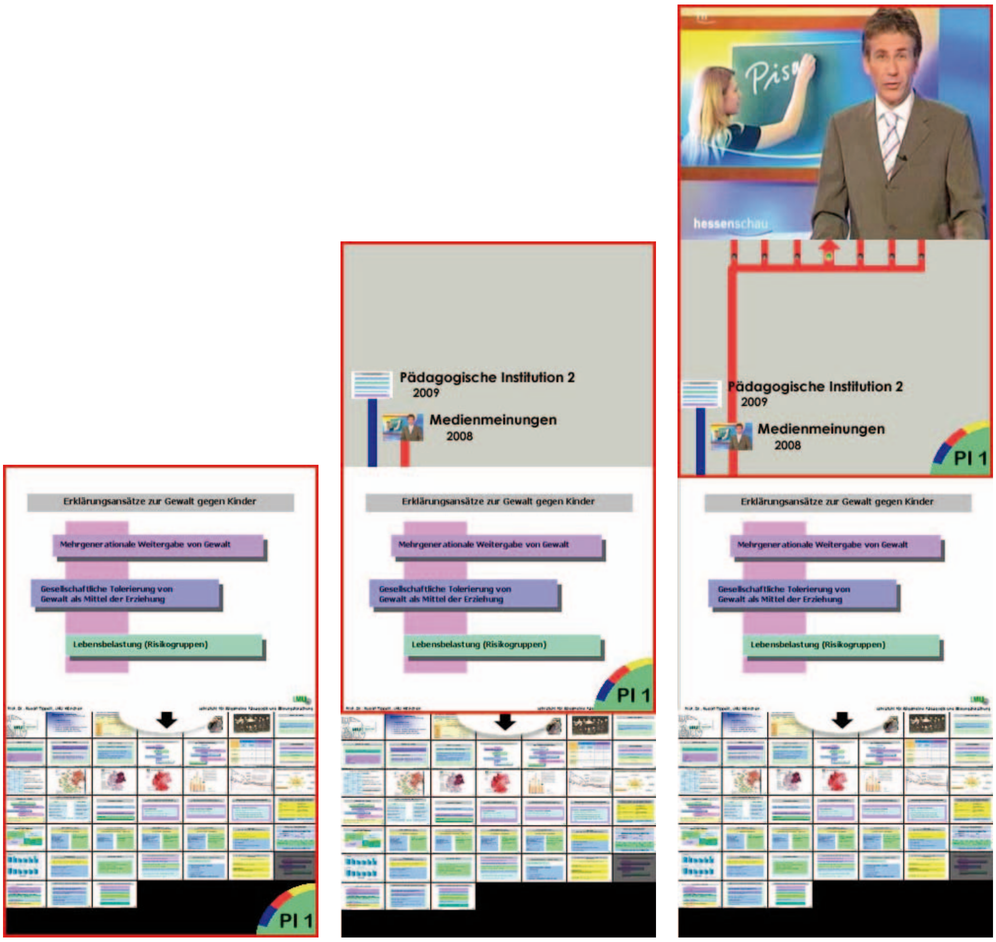


Figure 2.9: From left to right: navigation of topically inter-related videos. The red rectangle depicts the currently visible view. Here, the whole interface is visualized to illustrate the overall navigation history.

2.3.2.4 Results and Interim-discussion

In the following, the results from the data analysis are presented and discussed according to the level of navigation complexity.

Individual Segment Navigation. The participants stated that the granularity of the timeline seemed appropriate to navigate short videos of about 5 minutes in length. Focusing on the tilt-based concept, the participants raised concerns that the concept might not be adequate for mobile settings. Here, the concept might be too sensitive with respect to environmental noise such as accidental movements which therefore result in accidental navigation. The participants also mentioned that the point of view is crucial and since the device is being tilted, it is unclear whether they can preserve the full level of detail which is crucial for deciding when to stop forwarding or rewinding the video.

Large Video Navigation. Overall, the participants appreciated the two-dimensional concept, presenting four keyframes at a time. They argued that the keyframes are well laid out and this supports them in getting an overview. However, they reasoned that it might be difficult to grasp the context of the presented keyframes within the whole video. The grid-based layout was considered as clearly laid out and moreover, the participants believed that it would support their orientation within the video. However, they raised concerns with respect to zooming into the grid layout. One participant commented that only one zoom level might not be enough, while the other participants were content with having only one additional zoom level.

Inter-related Video Collection Navigation. The participants liked the idea of navigating between topically inter-related videos in the vertical dimension. However, they critiqued the visual appearance, as well as requiring the user to select related keyframes by tapping onto the forked lines in Figure 2.9 (right). The latter were considered as too "*small*" and "*error-prone*". Additionally, the arrow indicating related videos was considered misplaced, since it was hiding keyframes in the overview.

Figure 2.10: Lo-Fi prototype sketched by one participant.

Individual Segment Navigation	
✓	Granularity of timeline appropriate to navigate short videos
–	Tilt-based concept might be too sensitive for mobile settings
–	Tilting might constrain the point of view
Large Video Navigation	
✓	Grid-based layout supports preserving the orientation within a video
✓	2D concept supports getting an overview
–	2D concept makes it hard to grasp context of the presented keyframes
Inter-related Video Collection Navigation	
✓	Vertical dimension for topically inter-related videos appreciated
–	Forked lines too “small” and “error-prone”
–	Arrow indicating related videos cluttered the user interface
–	Color coding too “restrictive” and cognitively demanding

Table 2.1: Summary of the results from the first session.

Independent of the concrete navigation complexity, the participants commented on the color coding of different videos. They considered it to be “*too restrictive*”, since there might be a large amount of videos. These would then have to be mapped to a diverse set of colors, increasing the overall cognitive demand. Here, one participant proposed to use concrete descriptions of the video when touching the quarter circle (cf. Figure 2.10). The participant imagined that, once the quarter circle is touched, the available videos would be fanned out and presented alongside the quarter circle. As a matter of fact, the participants demanded further features such as bookmarking, which was not included in the initial concepts.

In summary, we decided to address the aforementioned issues (cf. Table 2.1) and to refine the concepts, discussing them in another focus group session. We particularly decided to

- drop the color coding of videos due to its complexity and
- include further concepts such as bookmarking to get an idea of how such functionality could be designed, eventually.

The concrete changes to the concepts are discussed in the following section.

2.3.3 Session 2: Prototype Refinement

The results from the first session showed that the interface concepts for both the navigation in large videos, as well as inter-related video collections were appreciated with respect to getting an overview over videos. However, the participants were mostly concerned with the corresponding visualization, as well as the proposed interaction techniques.

While the first paper prototypes were conceptually rather broad, the main objective of the second session was to unify the overall interface concept, focusing the concepts and therefore improving their stringency. The refined paper prototypes are discussed in the following.

2.3.3.1 Individual Segment Navigation

In the first session, the participants were mainly concerned about the tilt-based navigation. We have therefore applied only minor changes to the core interface.

We focussed on the remarks regarding the bookmark functionality and included it into this design iteration (cf. Figure 2.11). Now, tapping onto the display reveals the control elements, including an icon bar at the top. These icons can be dragged out by flicking from right to left. The second icon designates the access to the bookmark functionality. Playback can be paused, as well as resumed by tapping onto the button in the bottom right corner. Simply tapping onto the display once more hides the interface elements. The first icon allows a quick access to the slideshow interface for the large video navigation, discussed in the following subsection.

Figure 2.11: From left to right: Links to other views are visualized at the top of the screen once the user taps anywhere onto the interface. The bar at the top can be dragged out by flicking horizontally. The first icon designates a link to the slideshow interface for the navigation of large videos. The second icon (middle figure) shows a link to the bookmark functionality. Playback can be paused by tapping onto the icon in the right bottom corner.

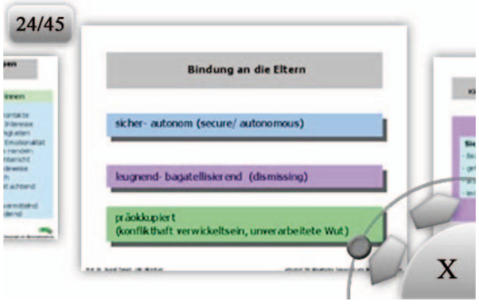


Figure 2.12: Slideshow navigation concept for large videos.

2.3.3.2 Large Video Navigation

We identified two major issues in the first session with respect to the navigation in larger videos: (1) the lack of context in the two-dimensional navigation, as well as (2) the applied color coding.

We have addressed the first issue by focusing on a one-dimensional keyframe layout in a focus+context interface (see Fig. 2.12). Here, the current keyframe is shown in the middle, whereas both pre- and proceeding keyframes are visualized to the left and to the right, respectively. Tapping onto the arrow keys at the bottom right allows to switch back and forth between the keyframes.

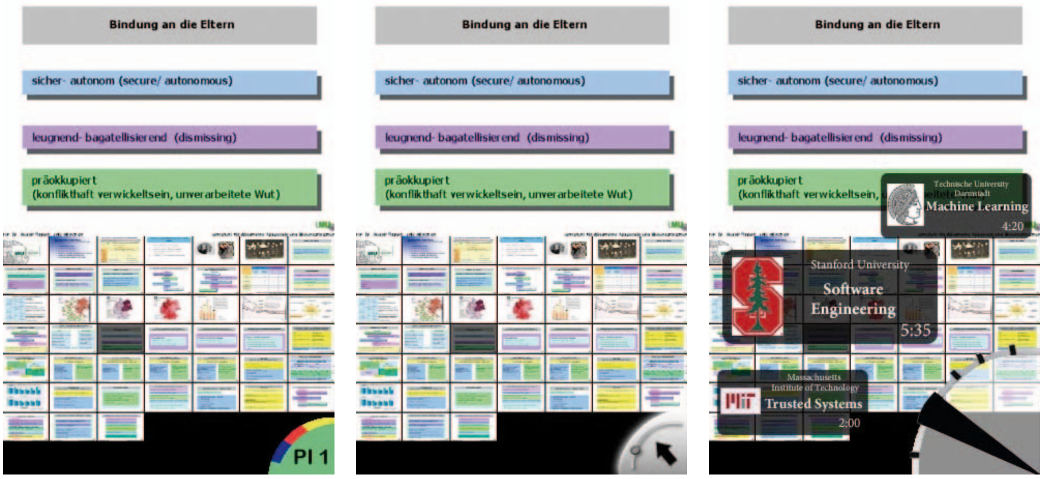


Figure 2.13: From left to right: The old overview+detail interface contained the color-coding for videos in the bottom right corner. This has been replaced in the next version by an arrow (middle figure), which expands to an overview over the available videos (right) by tapping onto it.

Inspired by the paper sketch in Figure 2.10, we have replaced the color coding with the button in the bottom right corner (cf. Fig. 2.13 left). By tapping onto the button as shown in Figure 2.13 (middle), other videos are revealed alongside the quarter circle (cf. Fig. 2.13 right).

2.3.3.3 Inter-related Video Collection Navigation

We have made only minor adjustments to the interface for the navigation in inter-related video collections: we have relocated the arrow indicating related videos to the top of the interface, to avoid any hiding of keyframes in the overview (cf. Figure 2.14). While the visual appearance of the inter-relations was criticized in the first session, we decided to keep it for a second session to to get a deeper understanding of how to design this particular sub-issue.

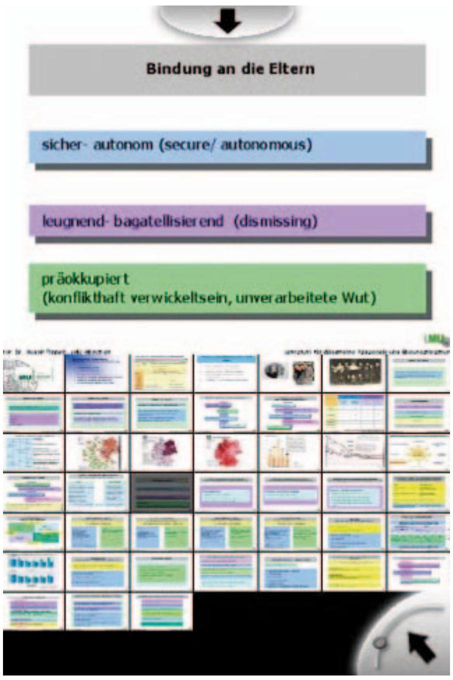


Figure 2.14: Improved navigation concept for inter-related video collections. The arrow indicating related videos is now located at the top of the interface. Related videos are revealed by dragging the arrow downwards.

2.3.3.4 Results

In the following, the results from the second session are presented according to the level of navigation complexity.

Individual Segment Navigation. the participants did not appreciate the shortcut bar at the top of the interface due to being "*too indirect*" and "*ine cient*". They stated that it contains too many functions and although bookmarking functionality might be helpful, they would sacrifice it for the sake of simplicity. They emphasized that the most important function for them is to switch to the slideshow interface, allowing them to quickly navigate larger videos. The participants suggested simple touch gestures for switching between modes.

Large Video Navigation. Overall, the participants appreciated the slideshow concept, as well as the grid-based overview+detail interface. With respect to the slideshow concept, they criticized the button-based interface. Comparable to the comments regarding the individual segment navigation, the participants found it to be too "*clumsy*" and "*indirect*". Although the buttons were "*conveniently located*", they were considered hard to touch. Thus, the participants suggested a more direct interaction, e.g. with the keyframes themselves. Regarding the overview+detail interface, the participants confirmed that one zoom level is appropriate, instead of introducing further levels of detail. Moreover, they disliked the newly introduced button to switch between videos and suggested to leverage simple gestures instead.

Inter-related Video Collection Navigation. In general, the participants again liked the idea of connecting topically related videos through vertical links. However, they criticized the applied GUI metaphor and argued that (1) the video thumbnails need to be larger, in order to be more easily recognizable and (2) the selection of related videos should be more intuitive; i.e. instead of tapping onto buttons located below a keyframes, they wanted to interact with the keyframes directly.

2.3.4 Summarizing Discussion

While the participants argued against including the bookmark functionality in the way we presented it, one of the participants proposed to add a bookmark

Figure 2.15: Lo-Fi prototype sketched by a participant during the second session. The proposed overview+detail interface was extended with a bookmarks bar in the middle. The participant imagined the bookmarks to be simple keyframe previews of the bookmarked videos as thumbnails.

bar to the grid-based overview+detail interface (see Figure 2.15).

We have explored various interface concepts during the two focus group sessions. Primarily, they were centered around the notion of navigation complexity for mobile video browsing. Within these distinct three levels of navigation complexity, we have introduced different concepts. Each of them leveraging the capabilities of different interface types, e.g. classical graphical user interface elements such as buttons, but also physical elements like the tilting of a device. Table 2.2 summarizes the main results from the second session and depicts general findings and implications which we will discuss in detail in the following.

The major requirement all of the participants shared across the sessions was that of direct interaction with the interface. During the first session, we relied on single-touch-based interface elements such as buttons, since they are well established through the classical WIMP paradigm. Although we addressed this in the second focus group session, the participants once more emphasized the importance by demanding e.g. direct interaction with thumbnails to browse through larger videos or direct access to related videos by interacting with the video keyframes in Figure 2.9.

General Findings / Design Implications	
✓	Direct interaction support for video playback
✓	Non-linear interaction support for larger videos and video collections
✓	Focus on core functionality and drop additional features (e.g. bookmarking)
Individual Segment Navigation	
✓	Simple touch gestures to switch between browsing modes
–	Shortcut bar “too indirect”
Large Video Navigation	
✓	Grid-based overview+detail interface preferred
–	Button-based interfaces considered “too indirect” and “clumsy”
Inter-related Video Collection Navigation	
–	Selection of keyframes too little intuitive → more direct interaction
–	Video thumbnails need to be easily recognizable

Table 2.2: Summary of the results from the second participatory design session.

The participants revealed various additional functional requirements such as bookmarking. The definite need for this could not be confirmed in the second session. The participants considered it as a helpful tool when re-using the video browser more often, but did not feel the need to include it as a basic functionality. Due to these reasons, we decided not to include it in our further designs and focus on the core requirements.

In conclusion, we set out to focus on two identified essential requirements for the final interface designs:

1. direct interaction support for video playback and
2. non-linear interaction support for both larger videos and inter-related video collections.

In addition to the three levels of navigation complexity, the explored interface types serve as a second dimension in the design space for mobile video browsing, which we discuss in the following section.

2.4 Design Space of Mobile Video Browsers

The iterative design process has highlighted crucial requirements for mobile video browsers, such as direct interaction support. This heavily depends on the utilized interface type. In addition to the three levels of navigation complexity (vertical axis in Figure 2.16), we regard the type of interaction used in the video browser as the second dimension in the design space. It is depicted as the horizontal axis in Fig. 2.16.

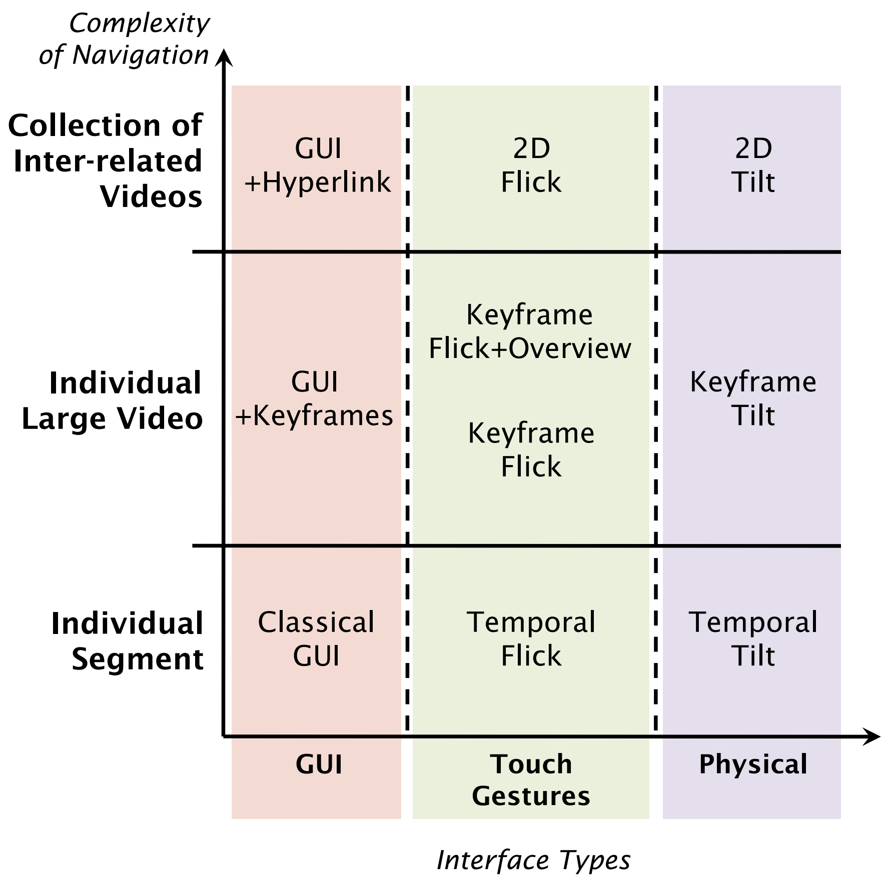


Figure 2.16: Design space for mobile video browsers.

These types range from classical interactions, which are well-known from desktop computing, to innovative interactions which leverage the additional capabilities of mobile devices, such as multi-touch displays and inertial sensors. Although the boundaries between the interface types are not selective, we regard them as discrete categories.

The categories comprise

- classical graphical user interfaces ported to the small display of mobile devices,
- interfaces that rely on gestures performed by touching the display of the mobile device and
- interfaces that rely on manipulating the device itself in the physical space.

The combination of both dimensions allowed us to identify the final interface concepts, which are presented in the following section.

2.5 Final Interface Concepts

So far, we have setup the design space for mobile video browsing according to two dimensions: the complexity of navigation and the type of interface the video browser relies on. We have situated one existing interface concept (the standard iPhone video player) and designed and implemented 7 novel interface concepts within the design space.

The interfaces addressing the navigation within a large video utilize keyframe abstraction [Truong and Venkatesh, 2007] to allow for an efficient overview over and a quick access to the contents of the video. The interface concepts for the navigation within a collection of inter-related videos are based upon hyperlinks. These hyperlinks exist between semantic segments of the video (e.g. key frames of a video or slides of a presentation recording), which can point to video segments or even a complete video. The hyperlinks are *1:n* relationships. Hence one semantic segment can point to various other semantic segments. It is out of the scope of this paper to investigate how these links can be created, since we focus on the navigation concepts. Hyperlinks could be created automatically through multimedia information retrieval [Rigamonti et al., 2007]. Furthermore, the interfaces could be enhanced to allow users to manually create (and share) links between videos.

2.5.1 GUI Navigation

A first class of interface concepts is based on the traditional GUI metaphor.

Classical GUI. The standard iPhone video player (see Fig. 2.17) is a classical GUI for the navigation within individual video segments. The interface elements in the center feature a play/pause button and two buttons to jump to the next (or previous) file or video in the playlist. The latter buttons can also be used for navigation (rewind or fast-forward respectively), when being pushed for a longer period of time. The slider below the buttons allows controlling the volume. The timeline at the top is used for the navigation by dragging the knob along it. However, the timeline itself cannot be manipulated. By tapping onto the knob a little longer and dragging it vertically activates a technique called scrubbing (comparable to the MobileZoomSlider [Sun and Hurst, 2008]): when a user drags the knob farther down vertically, the interface adapts the navigation granularity (in discrete levels). For instance by dragging the knob to the middle of the interface and then dragging it horizontally as if the user manipulated the timeline, the navigation speed is being reduced by 50% and the granularity is therefore increased.

GUI+Keyframes. Our GUI+Keyframes interface concept (see Fig. 2.18) is an enhanced version of the iPhone video player. We have added two buttons,

Figure 2.17: Classical GUI for an individual video segment.

Figure 2.18: GUI+Keyframes interface for an individual large video.

which allow switching back and forth between semantically segmented units and therefore navigating more quickly through a rather lengthy video. These buttons are located to the left (and right) of the previous (and next) buttons.

GUI+Hyperlink. As a further enhancement of the GUI+Keyframes interface concept our GUI+Hyperlink interface concept (see Fig. 2.19) supports the navigation between topically related videos, which are contained in an inter-related

Figure 2.19: GUI+Hyperlink interface for a collection of inter-related videos.

video collection. By tapping onto the screen, the user can access a list of hyperlinks relating the current segment to other videos. By tapping onto one of the links, the related video segment is replayed. The button located to the left of the list allows browsing back in the history, comparable to a back button in a web browser.

2.5.2 Gesture-based Navigation

The second class of interface concepts draws more extensively on the direct touch input capabilities of modern mobile devices. Instead of buttons (like in traditional GUIs), we leverage flick gestures that can be easily performed by touching the display. The gestures are inspired by the analogy to thumbing through a book.

Temporal Flick. Our Temporal Flick interface concept is shown in Figure 2.20. The timeline at the bottom of the interface is used for the navigation within an individual video. Additionally, users can navigate through the video's frames by flicking horizontally. Flicking from right to left fast-forwards and flicking from left to right rewinds, respectively. The playback speed depends on the length of the flick. Tapping the display toggles between the play and pause modes.

Keyframe Flick & Keyframe Flick+Overview. Two interfaces support the navigation within a large video. The Keyframe Flick interface (see Fig. 2.21) is an extension of the temporal flick interface. When the video is paused, the current key frame is displayed and by flicking horizontally, users can switch back and

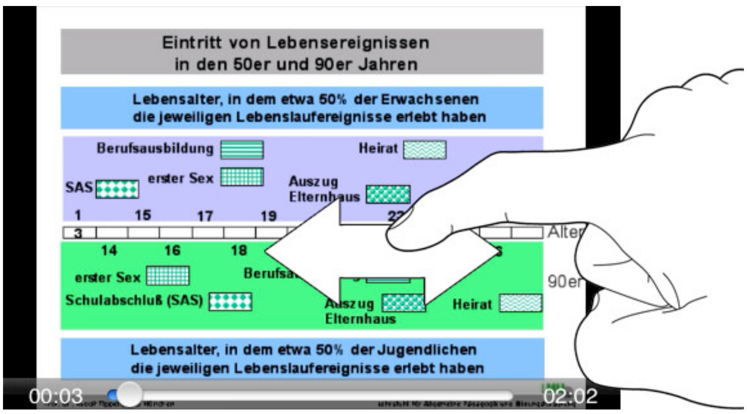


Figure 2.20: Temporal Flick interface.

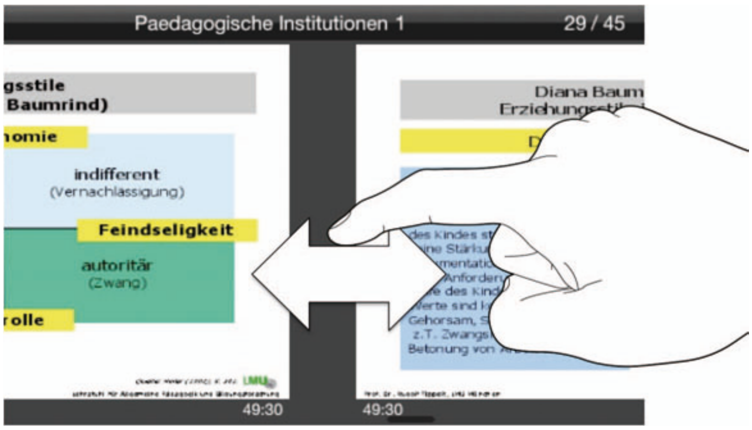


Figure 2.21: Keyframe Flick interface.

forth between key frames. By tapping the display again, playback of the video is resumed.

The Keyframe Flick+Overview interface (see Fig. 2.22) extends the keyframe flick interface by an overview with thumbnails of all keyframes of the video. These are displayed in the lower part of the interface in a grid layout. The currently active key frame is highlighted. In addition to the flick

Figure 2.22: Keyframe Flick+Overview interface.

interactions described above, the user can directly navigate to any key frame by tapping onto its thumbnail. Moreover, key frames can be skimmed very quickly by sliding the finger over the grid. Either rotating the device into landscape mode or double tapping on the current video in the upper part activates the keyframe flick interface.

2D Flick For the navigation between topically related videos, we contribute the 2D Flick interface concept (see Fig. 2.24). We aim at providing an intuitive interaction technique, which allows users to follow hyperlinks and navigate easily within the navigation history. The major challenge is to prevent users from getting lost in too much information presented on a small screen. Lost in Hypertext [Edwards and Hardman, 1999] is a well-known phenomenon, which may occur particularly in this situation. Therefore, we apply a spatial navigation concept (cf. Fig. 2.23): While segments within one single video can be accessed by flicking left and right (as described above), hyperlinks between different videos can be followed by flicking up and down.

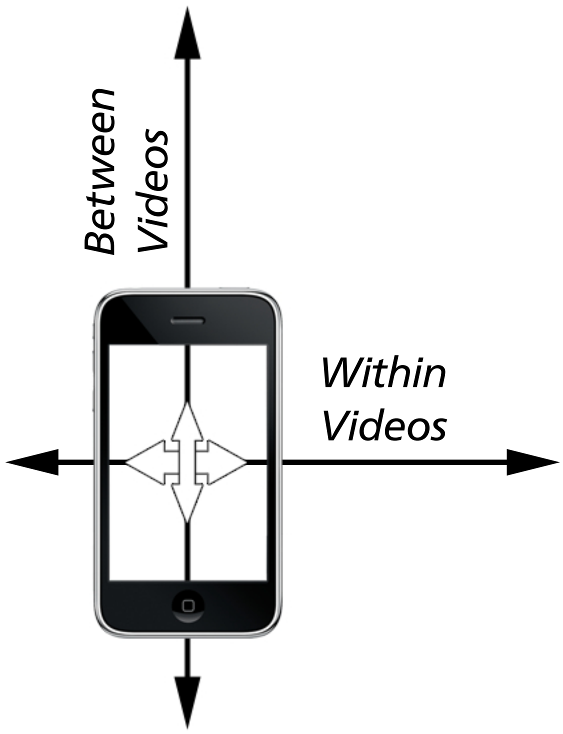


Figure 2.23: Spatial interaction concept: horizontal flicking browses within videos, vertical flicking browses between videos.

Figure 2.24: (a) 2D Flick interface, navigation between videos, (b) Visualized vertical browsing history.

Whenever a hyperlink is available, this is indicated by a small arrow in the upper right corner of the user interface (see Fig. 2.22 and 2.24 left). When the user flicks downwards, the interface is being scrolled downwards, revealing related videos as shown in the lower interface screenshot in Figure 2.24 (left). In this example, two interlinked videos (visualized using grey boxes) contain relevant material. By tapping on one of the videos, the interface is being scrolled down further, displaying the interlinked key frames of the related video (see the upper interface screenshot in Fig. 2.24 left). These can also contain topical relations to other videos, which are again visualized with a small arrow in the upper right corner.

By aligning semantically related videos vertically, the browsing history results in a vertical stack. This can be navigated by simply flicking vertically up and down respectively. Alternatively, to avoid repetitive flicking and to gain an overview on the browsing history, a visualization thereof can also be used for the vertical navigation as shown in Figure 2.24 (right). It is activated by tapping anywhere on the screen for more than one second and then displayed as an image on top of the current video. The visualization can be navigated by moving the finger vertically across the images.

2.5.3 Physical Navigation

A third class of interface concepts leverages the affordance of mobile devices to be manipulated in the physical space.

Temporal Tilt. Our Temporal Tilt user interface (see Fig. 2.25) contains only one visible interface element: the circular timeline at the bottom right. Tapping the display anywhere toggles between the play and pause modes. When tapping the display for a longer period of time, a slider appears (see top right in

Figure 2.25: Temporal Tilt interface: illustrating the interface elements.

Figure 2.26: Usage of the tilt interface, arrows indicate tilt directions.

Fig. 2.25 and 2.26 near thumb) that can be dragged to the right to activate the navigation mode. By tilting the device to the right, the user can fast-forward the video. By tilting to the left, the user can rewind the video. The interface features two discrete playback speeds, depending on the tilt angle (the greater, the faster or slower respectively). For detecting the tilt action, the device's accelerometer is used.

Keyframe Tilt & 2D Tilt. The same tilt-based concept can also be used for navigating between segments of large videos (Keyframe Tilt). Instead of fast-forwinding or rewinding the video, tilt actions result in jumping forth and back between segments. The 2D Tilt concept involves not only tilting to the left and to the right, but also tilting upwards and downwards for navigating within inter-related video collections, similarly to the 2D Flick interface. For our study we have only implemented and evaluated the temporal tilt interface for reasons that will be discussed in the next section.

2.6 Controlled Experiment

The concepts described above have been evaluated in a controlled experiment. We evaluated the usability (focusing on efficiency, effectiveness, learnability, as well as user satisfaction) and user experience of each user interface. Moreover, we analyzed the specific advantages and drawbacks of each interface type for mobile video browsing in the design space, depending on the navigation complexity. In the following subsections, we first describe our methodology and then report and discuss the results of the usability and user experience evaluation, as well as an in-depth usability error analysis.

2.6.1 Experiment Setup and Design

We have conducted a controlled experiment with 44 participants (30 male, 14 female) from different scientific backgrounds (i.a. mathematics, social sciences, medicine, pedagogy, physics and design). We chose a within-subject design. Each single-user session had a duration of 2 hours. The sessions were video-recorded, relevant task completion times were measured and semi-structured interviews were conducted. Additionally, quantitative feedback was gathered using the well-known standard usability scale (SUS) questionnaire [Brooke, 1996]. The attractiveness of each user interface has been assessed using the AttrakDiff questionnaire [Hassenzahl, Platz, Burmester, and Lehner, 2000]. Our analysis comprises statistical measures and a detailed ascriptive analysis of 18 hours of video recordings.

2.6.2 Methodology

The tasks of the participants comprised simple fact finding and more complex knowledge integration tasks. For each task, a different data set was utilized to exclude any learning effects. Moreover, the order in which the interfaces were presented to the participants was counter-balanced. The tasks are outlined in the following.

2.6.2.1 Navigation in an Individual Segment

In order to evaluate and to compare the interface concepts for navigation within an individual segment, the participants were asked to perform two different fact-finding tasks with each user interface. As data, we utilized videos of about 5 minutes length. The first task required textual orientation, whereas the second task focused on visual orientation, since the user's orientation is crucial to quickly retrieve a desired part of a video. The participants had to fulfill the following tasks:

- **Task 1:** The participants were asked to search a short video for a certain topic. The position within the video was not revealed to them beforehand.
- **Task 2:** The participants had to find a specific scene in the video. They were shown a distinctive key frame of the scene beforehand.

2.6.2.2 Navigation in a Large Video

To assess the navigation in a large video, the participants were asked to complete three different fact-finding tasks. As data, we used lecture recordings of each about 90 minutes length and the corresponding slides as key frames. The tasks required visual orientation within a video (task 1 and 3), as well as textual orientation (task 2):

- **Task 1:** The participants had to search a given slide within a lecture recording without prior knowledge of the lecture.
- **Task 2:** The participants were asked to find a certain topic. They were advised of the fact, that it was contained in the last third of the lecture.
- **Task 3:** The participants had to navigate to the slide which directly follows the one found in the first task.

2.6.2.3 Navigation between Inter-related Videos

We also assessed the navigation in a collection of inter-related videos. The collection consisted of 7 lecture recordings (each about 90 minutes) and 6 news broadcasts (each about 15-30 minutes). We segmented the videos and manually related the segments topically. The participants had to fulfill the following tasks:

- **Task 1:** The participants were asked to complete a complex visual and textual fact-finding task involving multiple videos using both interfaces.
- **Task 2:** The participants had to complete a knowledge integration task for a given topic covered in multiple videos.

In Task 2, we used the same data set for both interfaces. To exclude any learning effects, we used a between-subject design for this particular task.

2.6.3 Results I: Usability

2.6.3.1 Navigation in an Individual Segment

Figure 2.27 shows an overview of the average time required for performing the tasks with each interface. Although the participants performed the tasks faster using the temporal flick interface, the one-way repeated measures ANOVA test revealed that the speed-up was not significant (see Table 2.3). The SUS score for the classical GUI browser is 76.53 (SD=15.01), 74.89 (SD=16.91) for the temporal flick browser and 62.44 (SD=16.87) for the temporal tilt browser. Hence, both the classical GUI browser and the temporal flick browser were perceived as most usable.

Overall, temporal flick turned out to be the best technique for temporal navigation within individual video segments. With the temporal tilt interface, users performed faster or not slower than with the classical GUI. We introduced the scrubbing feature of the classical GUI browser to our participants. However, none of the participants actually used this feature while browsing an individual segment.

Our early focus group studies (as discussed in Section 2.3), the qualitative analysis of the video data and statements from the semi-structured interviews all show that the tilt interaction is not adequate for continuously browsing within a video. Due to the larger viewing angle caused by tilting, the video on the display was less well viewable. Users felt that they need better visibility

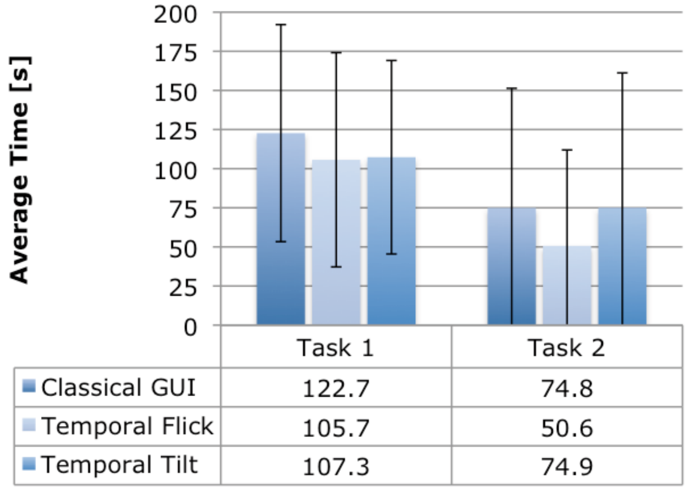


Figure 2.27: Average times for navigation in an individual segment (here and in the following error bars indicate standard deviation).

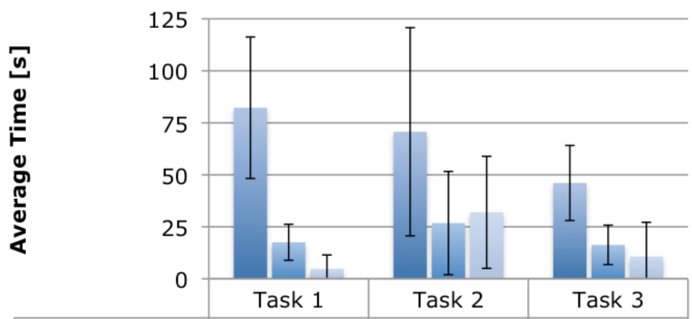
of the video for fine control in browsing. Moreover, users complained about the lack of haptical feedback. Vibro-tactile feedback was not an option at the point of the study, since the iPhone as our implementation platform supports only one constant vibration force. Future work should examine more deeply how haptic feedback can improve physical interfaces for mobile video browsing. Because of these negative results even in this rather simple setting of navigation within an individual video segment, we have opted for not evaluating this concept for the more complex tasks.

An interesting observation we have made was that in spite of the rather continuous interaction techniques like flicking or tilting, which require continuous motion, users performed compound interactions, consisting of several discrete, additive flicks or tilts, e.g. flicking once for navigating 5 seconds forward, twice for 10 seconds, three times for 15 seconds and so on. Users also transferred interactions from everyday actions to the physical interface types. One particular user for instance drew circles onto the display using the temporal flick interface. He wanted to replay a certain scene in a loop. This underlines the potential of these novel interaction techniques for mobile video browsing.

2.6.3.2 Navigation in a Large Video

Figure 2.28 shows an overview of the average time required for performing the navigation tasks in a large video with each interface. The participants were able to complete all three tasks significantly faster using either the keyframe flick or the keyframe flick+overview interface than using the GUI+keyframe interface (see Table 2.3 for the ANOVA results and Table 2.4 for the Bonferroni post-hoc test results). Comparing the keyframe flick with the keyframe flick+overview interface, we found that the participants were significantly faster using the keyframe flick+overview interface for task 1 and task 3. The difference in task 2 was not significant. This is in line with qualitative findings from the semi-structured interviews. The participants stated that the keyframe flick+overview interface supports their visual orientation and navigation (as in task 1 and 3), whereas they prefer to skim through the slides by flicking horizontally when they have no visual clues (as in task 2).

The SUS score is 89.26 (SD=9.2) for the keyframe flick+overview interface, 90.34 (SD=7.48) for the keyframe flick interface and 61.59 (SD=16.1) for the GUI+keyframe interface. Hence, both keyframe flick and keyframe flick+overview interfaces were perceived as far more usable than the GUI+keyframe interface.



	Task 1	Task 2	Task 3
GUI+Keyframe	82.3	70.7	46.1
Keyframe Flick	17.6	26.8	16.3
Keyframe Flick +Overview	4.8	32.0	10.8

Figure 2.28: Average times for navigation in a large video.

Navigation Complexity	Task	F	df	Sig.
Individual Segment	1	0.99	2, 86	> 0.05
	2	1.30	2, 86	> 0.05
Large Video	1	160.88	1.14, 49.09	< 0.001
	2	20.51	1.66, 71.37	< 0.001
	3	70.79	2, 86	< 0.001

Table 2.3: ANOVA results for the navigation time in an individual segment and a large video.

2.6.3.3 Navigation between Inter-related Videos

In both tasks, a t-test (repeated measures for task 1 and independent measures for task 2 respectively) showed that the participants were significantly faster ($p < 0.001$) using the 2D flick interface as shown in Figure 2.29. Moreover, statements in the interviews showed that the two dimensional browsing metaphor fosters the users' awareness of interrelated videos. Together, the above results show that a two-dimensional navigation metaphor supports the user's orientation when navigating across multiple videos.

The SUS score for the 2D flick interface is 83.07 (SD=12.33), whereas the GUI+Hyperlink interface scored 58.98 (SD=19.76). Consequently, the 2D flick

Task	Interface A	Interface B	$CI_{.999}$ (lower)	$CI_{.999}$ (upper)	Sig.
1	GUI+keyfr.	Keyfr. flick	42.85	86.51	< 0.001
	GUI+keyfr.	Keyfr. f+o	56.04	98.82	< 0.001
	Keyfr. flick	Keyfr. f+o	6.156	19.35	< 0.001
2	GUI+keyfr.	Keyfr. flick	10.37	77.45	< 0.001
	GUI+keyfr.	Keyfr. f+o	7.13	70.37	< 0.001
	Keyfr. flick	Keyfr. f+o	-26.08	15.77	> 0.05
3	GUI+keyfr.	Keyfr. flick	17.00	42.54	< 0.001
	GUI+keyfr.	Keyfr. f+o	21.34	49.34	< 0.001
	Keyfr. flick	Keyfr. f+o	-4.74	15.88	< 0.001

Table 2.4: Bonferroni test for the navigation in a large video.

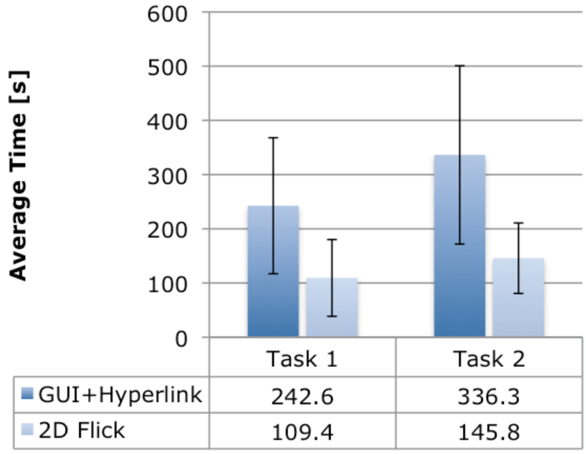


Figure 2.29: Average times for navigation in a collection of inter-related videos.

interface was perceived as far more usable than the GUI+Hyperlink interface. In the interviews, the participants commented on the spatial concept as "clearly laid out" and they remarked that the vertical alignment of the related videos intensifies the visual relationship between the videos.

2.6.4 Results II: User Experience

In addition to the usability evaluation described in the previous section, we have also measured the user experience for each interface. For this purpose, we have used the AttrakDiff questionnaire, which measures the attractiveness of interactive products and interfaces in particular. The questionnaire is comprised of a set of bi-polar attributes, characterizing a user's perception of the interface.

As stated by Hassenzahl, Burmester, and Koller [2003], AttrakDiff evaluates the following user experience dimensions:

Pragmatic Quality (PQ): Describes the usability of a product and indicates how successfully users are in achieving their goals using the product.

Hedonic Quality - Identity (HQ-I): Indicates to what extent the product allows the user to identify with it.

Hedonic quality - Stimulation (HQ-S): Mankind has an inherent need to develop and move forward. This dimension indicates to what extent the product can support those needs in terms of novel, interesting, and stimulating functions, contents, and interaction- and presentation-styles.

- **Attractiveness (ATT):** Describes a global value of the product based on the quality perception.

Hedonic and pragmatic qualities are independent of one another, and contribute equally to the rating of attractiveness [Hassenzahl et al., 2003].

In the following, the results of the user experience evaluation are presented according to the navigation complexity.

2.6.4.1 Navigation in an Individual Segment

Figures 2.30 and 2.31 illustrate the average scores of the bi-polar attributes for the classical GUI, temporal flick, as well as temporal tilt interfaces. Both temporal flick and temporal tilt interfaces excel in the hedonic dimensions. Here, the classical GUI interface shows a negative tendency, being characterized as rather *conventional*, *conservative* and *ordinary*. Even more so, the temporal tilt interface has the most positive scores with respect to the overall hedonic qualities. However, the temporal tilt interface has weak pragmatic qualities, being described as *cumbersome* and *unruly*. This is inline with the results from the usability evaluation. This phenomenon is further explained in the usability error analysis (see next subsection).

The overall attractiveness is visualized in a portfolio shown in Figure 2.32. The portfolio depicts the average hedonic (vertical axis), as well as pragmatic qualities (horizontal axis) for each interface. The ratings are categorized into 9 different "character regions", with "desired" depicting a high characteristic in both quality dimensions. Moreover, each interface is shown within a confidence rectangle, explaining whether the participants were at one with their rating. Thus a small confidence rectangle is to be regarded as positive, meaning less coincidental results. In Figure 2.32, all interfaces have a small confidence rectangle. Largely, the interfaces have been assessed as being *neutral*. Meaning that the interfaces can be improved in terms of their hedonic, as well as pragmatic qualities. Interestingly, the temporal tilt interface has a slight tendency toward being described as *self-oriented*, which attributes to its higher hedonic qualities. The classical GUI interface is to be considered the least attractive with the least characteristic in both dimensions, although both classical GUI and temporal flick are comparable in terms of their pragmatic quality.

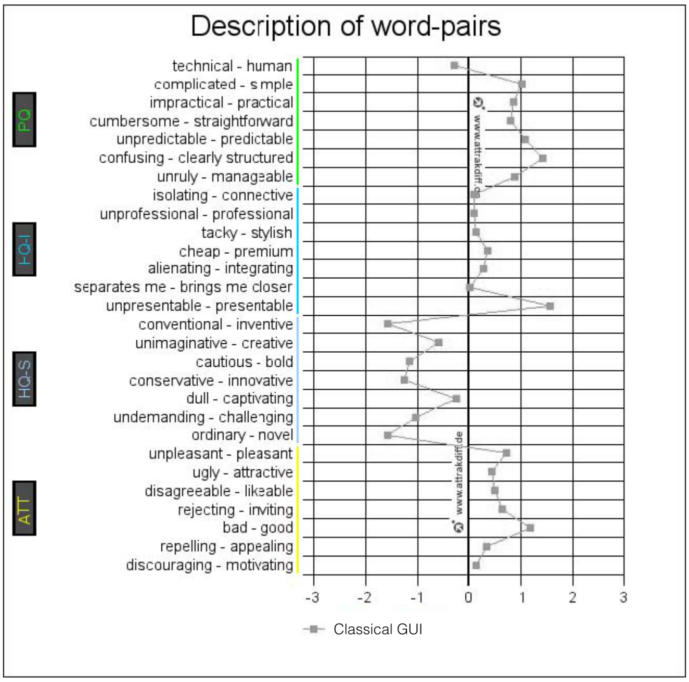


Figure 2.30: Description of word-pairs for the classical GUI interface.

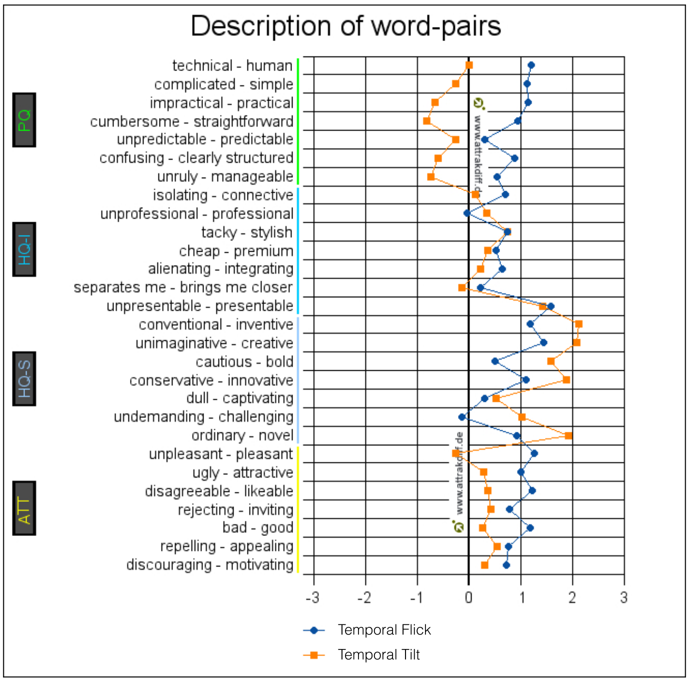


Figure 2.31: Description of word-pairs of both temporal flick and temporal tilt interfaces.

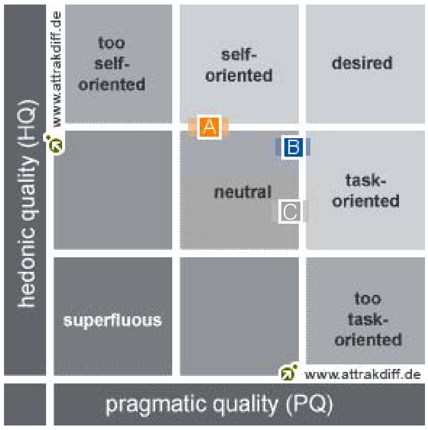


Figure 2.32: Attractiveness portfolio for the interfaces supporting the navigation in individual video segments. (A) Temporal tilt, (B) temporal flick, (C) classical GUI.

2.6.4.2 Navigation in a Large Video

The keyframe flick+overview interface has achieved better average scores than the keyframe flick interface. It particularly excels in both pragmatic qualities, as well as hedonic qualities with respect to stimulation. Comparing these two interfaces, the flick+overview interfaces was only considered *more technical*, *less*

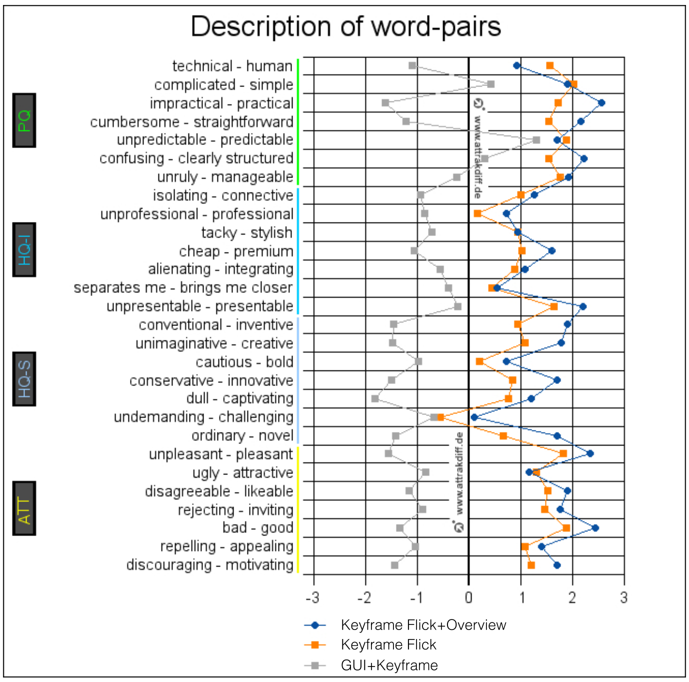


Figure 2.33: Description of word-pairs of the interfaces for the navigation in large videos.

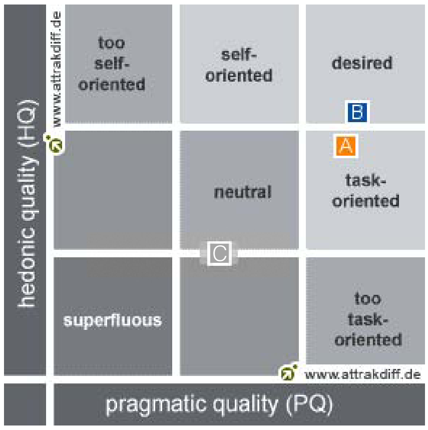


Figure 2.34: Attractiveness portfolio for the interfaces supporting the navigation in large videos. (A) Keyframe Flick, (B) Keyframe Flick+Overview, (C) GUI+Keyframe.

predictable and *more ugly*. Figure 2.33 shows the average scores of the bi-polar attributes of the interfaces for the navigation in large videos. The flick-based interfaces, keyframe flick and keyframe flick+overview, excel in all dimensions. Although the GUI+keyframe interface is described as rather *simple* and foremost *predictable*, it is inferior to the flick-based interfaces in terms of its hedonic qualities and consequently in its overall attractiveness. The GUI+keyframe interface is regarded as being *cumbersome*, *conventional*, *dull* and *unpleasant*.

The attractiveness portfolio in Figure 2.34 further underlines these results. The GUI+keyframe interface is assessed as *neutral* with a slight tendency toward being *super uous*, underlining its mediocre values in both quality dimensions. The keyframe flick interface is assessed as being *task-oriented* with an average characteristic in the hedonic quality dimension. Furthermore, the keyframe flick+overview interface is the most attractive interface, being described as *desired*. Both keyframe flick and keyframe flick+overview interfaces have minimal confidence rectangles.

2.6.4.3 Navigation between Inter-related Videos

The results of the AttrakDiff evaluation for the interfaces supporting the navigation in large inter-related video collections are shown in Figure 2.35. The description of word-pairs shows that the 2D flick interface excels in all dimensions, with being i.a. perceived as *practical*, *inventive*, *innovative*, *novel*, *good*, as well as *motivating*. The GUI+Hyperlink interface has achieved only mediocre results with negative tendencies in its pragmatic qualities. Moreover, it was i.a. perceived as *technical*, *dull* and *discouraging*. The attractiveness portfolio in Figure 2.36 illustrates these results. While the GUI+Hyperlink interface is categorized as "neutral". The 2D flick interface is situated within the "desired" region, attributing to its strong characteristics in both quality dimensions.

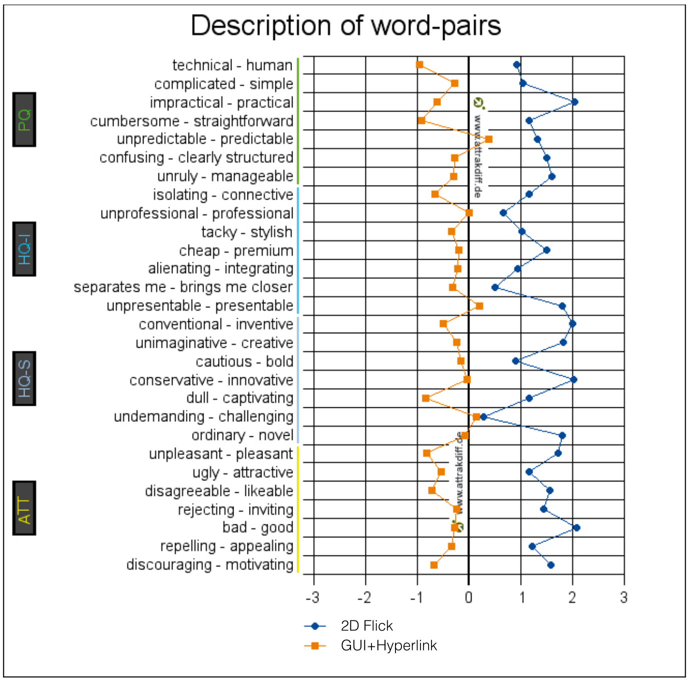


Figure 2.35: Description of word-pairs of the interfaces for the navigation in inter-related video collections.

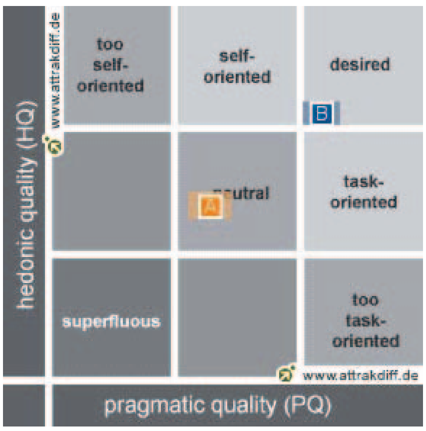


Figure 2.36: Attractiveness portfolio for the interfaces supporting the navigation in inter-related video collections. (A) GUI+Hyperlink, (B) 2D Flick.

2.6.5 Usability Error Analysis

For a deep understanding of the problems related to specific interface concepts, it is important to examine which errors are made when using the interfaces. We therefore performed a detailed analysis of usability errors which occurred during our experiment. From the detailed descriptions of individual errors, we derived a general error taxonomy for mobile video browsers. In the following subsections, we first report on our methodology, present our error classification and report the results of our error analysis. Together with our findings from the previous section, these provide the basis for design implications.

2.6.5.1 Methodology

Inspired by Masson, Hill, Conner, and Guidon [1988], we utilize ascription for the analysis of the video data collected in the controlled experiment. Since we recorded the interactions from behind the participants' shoulders, the interfaces were always clearly visible. We have coded potential errors using a template describing the type of error, a detailed description of the error, its impact on the efficiency (e.g. loss of internal locus of control) and the occurrences per task and user interface. In the following, we outline the classes of our error taxonomy and then present the results of our analysis.

2.6.5.2 Error Classes

We have identified the following four abstract error classes:

- **E1, Interface element not accessible:** The interface element could not be manipulated by the user. Typical reasons for this error are small or misplaced interface elements.

- **E2, Interface element was used incorrectly:** This type of error designates incorrect interactions. Common errors of this type are for instance wrong gestures.

- **E3, Interface elements misinterpreted:** Errors of this type mostly happened due to misconceptions. For instance, interface elements of the same type (e.g. two sliders), which were mapped onto different functions, were confused.

- **Slips:** All other errors, which do not belong to any of the above classes, are called slips (in the sense of [Norman, 2002]), e.g. users performed certain actions accidentally.

The interface-specific errors within these categories are discussed in the following subsections.

2.6.5.3 Navigation in an Individual Segment

Although the GUI concept is the most well known interface concept in our design space, the participants committed most of the errors using the standard iPhone video player. They committed significantly less errors using both the temporal flick (73% less) and tilt browser (79% less). Other differences were not significant (cf. Table 2.6 for the ANOVA results and Table 2.7 for the Bonferroni post hoc test results).

	Class. GUI		Temp. Flick		Temp. Tilt	
Task	1	2	1	2	1	2
E1	**71**	**58**	**14**	4	0	0
E2	15	4	3	0	**13**	3
E3	5	4	5	10	7	6
Slip	31	34	11	**11**	11	**6**
Sum	122	100	33	25	31	15

Table 2.5: Amount of errors for the navigation in an individual segment (here and in the following, bold numbers indicate the peak per task).

The majority of the errors made with the classical GUI concept (the standard iPhone video player) were of type E1 (see Table 2.5). Users were unable to navigate through the video using the timeline placed at the top of the interface. Placing the timeline at the top of the interface causes severe issues. Most commonly, a mobile device is used in landscape mode to browse a video, since it offers the most screen real estate. In our experiment, a significant amount of participants held the device in both hands, such that only the thumbs are able to interact with the interface. The rest of the hand is located behind the device. Consequently, the interaction is highly limited by the length of the

Navigation Complexity	Task	F	df	Sig.
Individual Segment	1	7.15	1.11, 47.90	< 0.01
	2	5.14	1.25, 53.82	< 0.05
Large Video	1	9.11	1.27, 54.77	< 0.01
	2	2.72	1.27, 54.74	< 0.05
	3	4.24	1.05, 44.93	< 0.05

Table 2.6: ANOVA results for the errors during the navigation in an individual segment and a large video.

users' thumbs. Figure 2.37 shows one of our participants while trying to use his thumb to interact with the timeline of the iPhone video player. Since the timeline is located at the top of the interface and is not in reach for his thumb, he needs to lift his right hand, therefore occluding nearly the entire display real estate. Another example is shown in Figure 2.38 where a participant tries to grab the timeline of the GUI interface up-side-down with his thumb, leading to a usability error, since he could not grab the knob appropriately. In the case of the GUI interface, the timeline should be placed at the bottom of the interface to (1) minimize the navigation paths and (2) prevent users from occluding the screen while using the timeline for navigation.

Slips were also a severe problem in case of the standard iPhone video player. Users often tapped onto the "next title" button accidentally and therefore stopped the playback of the current video. Consequently, they lost the in-

Task	Interface A	Interface B	$CI_{.95}$ (lower)	$CI_{.95}$ (upper)	Sig.
1	Class. GUI	Temp. Flick	0.10	4.03	< 0.05
	Class. GUI	Temp. Tilt	0.26	4.02	< 0.05
	Temp. Flick	Temp. Tilt	-0.47	0.60	> 0.05
2	Class. GUI	Temp. Flick	0.12	3.75	< 0.05
	Class. GUI	Temp. Tilt	-0.35	3.58	< 0.05
	Temp. Flick	Temp. Tilt	-1.10	0.47	> 0.05

Table 2.7: Bonferroni test results for the errors during the navigation in an individual segment.

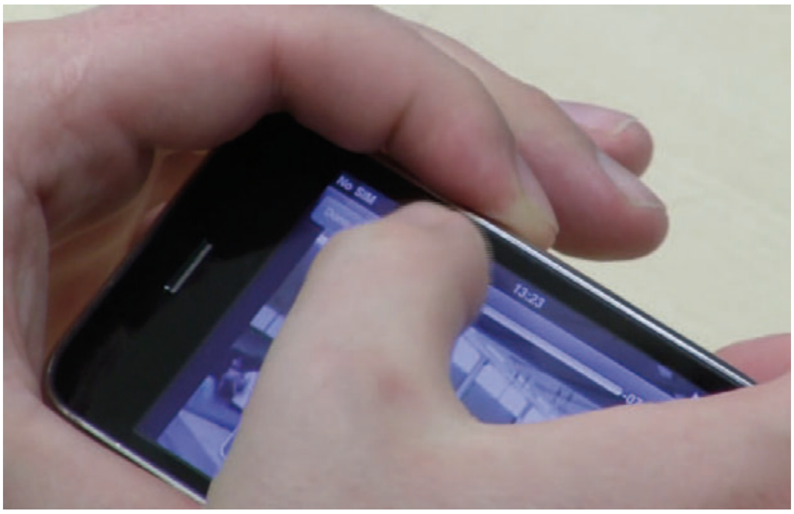


Figure 2.37: Timeline of the GUI interface is difficult to reach using a thumb, since it is placed at top.

ternal locus of control, had to restart the video and continue their video search from the beginning. Another difficulty with the iPhone video player was the fact that the same slider interface element was used for both timeline and volume control. Both interface elements got confused frequently (see E2 in Table 2.5). Slips were the most dominant error type for both temporal flick and tilt interfaces. However, although they committed only little slips, the amount can be further reduced when users become more familiar with such novel inter-

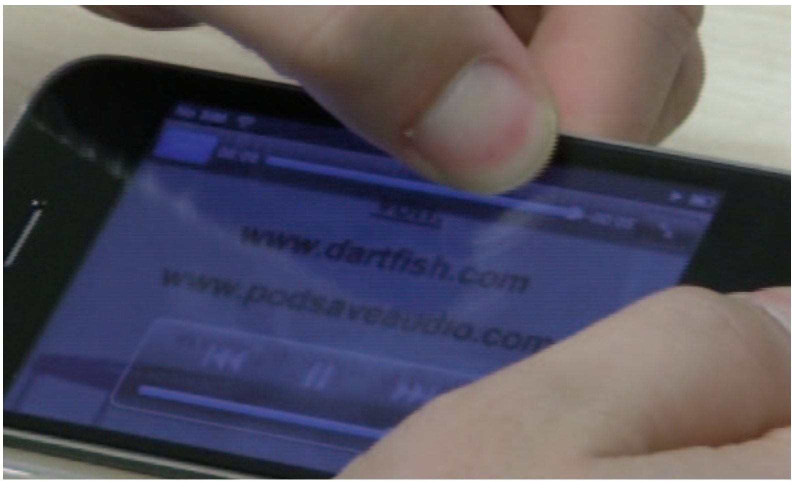


Figure 2.38: A participant tries to grab the timeline of the GUI interface up-side-down with his thumb.

action techniques. Regarding the temporal tilt interface, users had difficulties with enabling the navigation mode (error E2 in Table 2.5). Moreover, errors of type E3 were also problematic for both interfaces. With both interfaces it occurred that participants confused the correct navigation directions, e.g. flicking from left to right to navigate forward, instead of flicking from right to left. This is possibly due to differently remembered experiences and therefore a different mental model of the interface.

2.6.5.4 Navigation in a Large Video

Table 2.8 shows an overview of the amount of errors committed with each user interface. Virtually no errors were made with the gesture-based interfaces due to their high usability. Again, the ANOVA test (cf. Table 2.6 and Table 2.9 for the Bonferroni post-hoc test results) showed, that the participants committed significantly less errors using either the keyframe flick (73% less) or keyframe flick+overview interface (91% less) than using the GUI+keyframe interface. Most errors using the GUI+keyframe interface were again of type E1 (see Table 2.8). Users also had the problem of dealing with the timeline at the top of the interface. In case of the keyframe flick interface, the few errors resulted from flicking too hesitantly. The same also holds for the keyframe flick+overview interface.

	GUI+Keyframe			Keyframe Flick			Keyframe F+O		
Task	1	2	3	1	2	3	1	2	3
E1	**22**	**32**	**54**	0	0	0	0	0	0
E2	6	8	7	**10**	**9**	5	0	6	0
E3	4	3	1	0	0	0	0	0	0
Slip	6	10	10	3	7	**10**	**1**	**6**	**1**
Sum	38	53	72	13	16	15	1	12	1

Table 2.8: Amount of errors for the navigation in a large video.

Task	Interface A	Interface B	CI.95 (lower)	CI.95 (upper)	Sig.
1	GUI+keyfr.	Keyfr. flick	-0.18	1.15	< 0.05
	GUI+keyfr.	Keyfr. f+o	0.25	1.43	< 0.01
	Keyfr. flick	Keyfr. f+o	0.03	0.52	< 0.05
2	GUI+keyfr.	Keyfr. flick	-0.45	2.13	< 0.05
	GUI+keyfr.	Keyfr. f+o	-0.35	2.21	< 0.05
	Keyfr. flick	Keyfr. f+o	-0.45	0.63	> 0.05
3	GUI+keyfr.	Keyfr. flick	-0.51	3.10	< 0.05
	GUI+keyfr.	Keyfr. f+o	-0.14	3.36	< 0.05
	Keyfr. flick	Keyfr. f+o	0.01	0.63	< 0.05

Table 2.9: Bonferroni test results for the errors during the navigation in a large video.

2.6.5.5 Navigation between Inter-related Videos

An overview on the amount of errors of the navigation between inter-related videos is given in Table 2.10. T-tests showed that the 2D flick interface concept was significantly less error prone (81% less) than the GUI+hyperlink interface for both tasks ($p < 0.001$). Again, this is also due to the high usability of the gesture-based interface. The most common errors for the GUI+hyperlink interface were again of type E1 (see Table 2.10), due to the misplaced timeline. The errors with the 2d flick interface were mostly slips.

	GUI+Hyperlink		2D Flick	
Task	1	2	1	2
E1	**42**	**26**	0	0
E2	19	9	**10**	0
E3	2	1	0	0
Slip	17	4	8	**5**
Sum	80	40	18	5

Table 2.10: Amount of errors for the navigation in a collection of inter-related videos.

2.7 Orthogonal Design Principles

The evaluation results are summarized in Figure 2.39. The touch-based interfaces excelled in both their usability and their attractiveness, whereas the GUI-based interfaces were perceived as least usable and least attractive. The physical interfaces had high hedonic but only little pragmatic qualities. Based on the analysis of the different interface concepts within the design space, we derive principles for the design of mobile video browsers in the following.

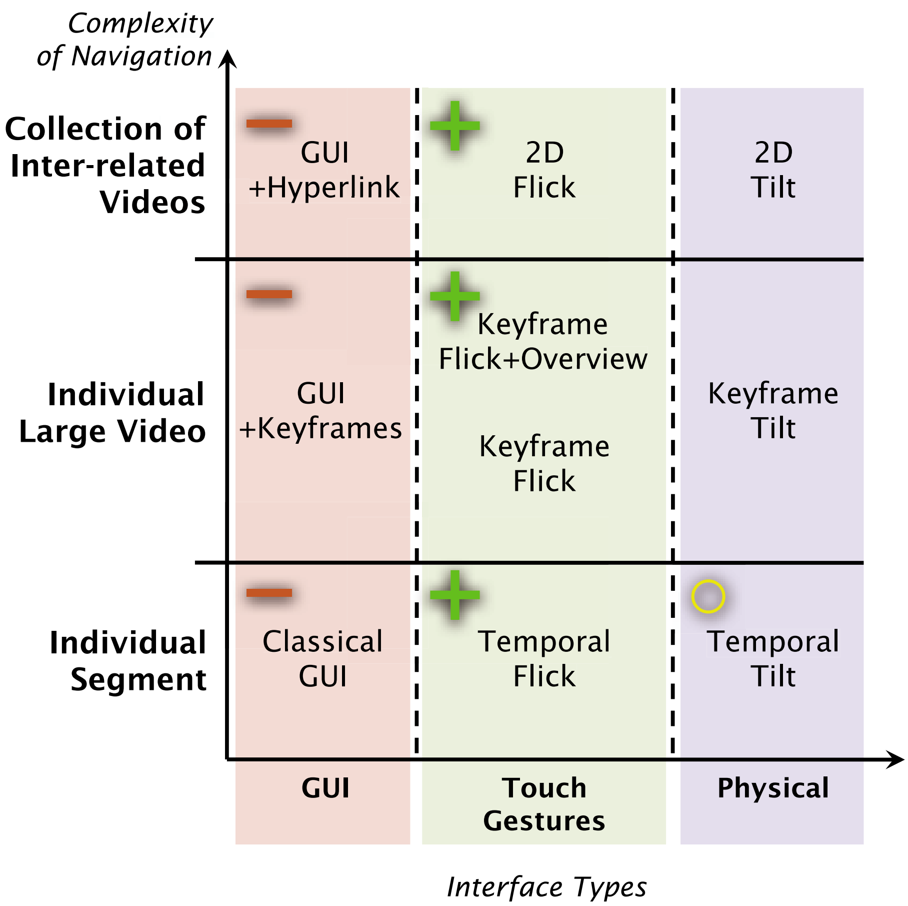


Figure 2.39: Summary of the evaluation results. The icons correspond to the overall ratings. The GUI-based interfaces were inferior in terms of both usability and user experience. The physical interfaces were considered most attractive, but least usable. The gesture-based interfaces excelled in both dimensions.

2.7.1 Spatio-temporal Browsing with Flick Interactions

The evaluation showed that flick gestures are a highly efficient concept for temporal navigation within short video segments. Not only users navigated more quickly with this interface than with a time-slider or the tilt-based interaction concept. Moreover, the number of errors made with the flick interface was significantly lower than with the GUI-based interfaces. This holds for all complexity levels, as illustrated in Figure 2.39. The evaluation also shows that by applying a spatial interaction metaphor in combination with simple, but highly efficient flick gestures, users are able to build up a mental model even of highly complex information spaces like inter-related digital libraries. Future interfaces should leverage spatial interaction concepts with flick interactions to support users effectively, instead of relying on hypertext-inspired navigation concepts (cf. 2D Flick versus GUI+Hyperlink, Fig. 2.39).

2.7.2 Support for Discrete Temporal Navigation

Although concepts for temporal navigation may afford continuous manipulation (e.g. continuously wiping over the display and therefore navigating through a video continuously), our observations have shown that users demand the possibility to navigate in discrete steps. By performing one gesture (e.g. flick or tilt) the video should be moved forward or backward by a fixed amount of time, for instance 10 seconds. These gestures should be additive: By repeating the gesture several times, a larger amount of time can be navigated forward (or backward respectively). Compared to a continuous interaction, where the video is being navigated as long as the gesture is performed, this has the advantage to offer better control and reversibility of the command. In particular, this is not possible with current timeline-based interfaces, such as the standard iPhone movie player (cf. classical GUI interface in Figure 2.39).

2.7.3 Place GUI Elements to Be Reachable by the User's Thumb

Classical GUI elements should be reachable by a user's thumbs. This guideline might appear straightforward. However, the evaluation shows that wide-spread

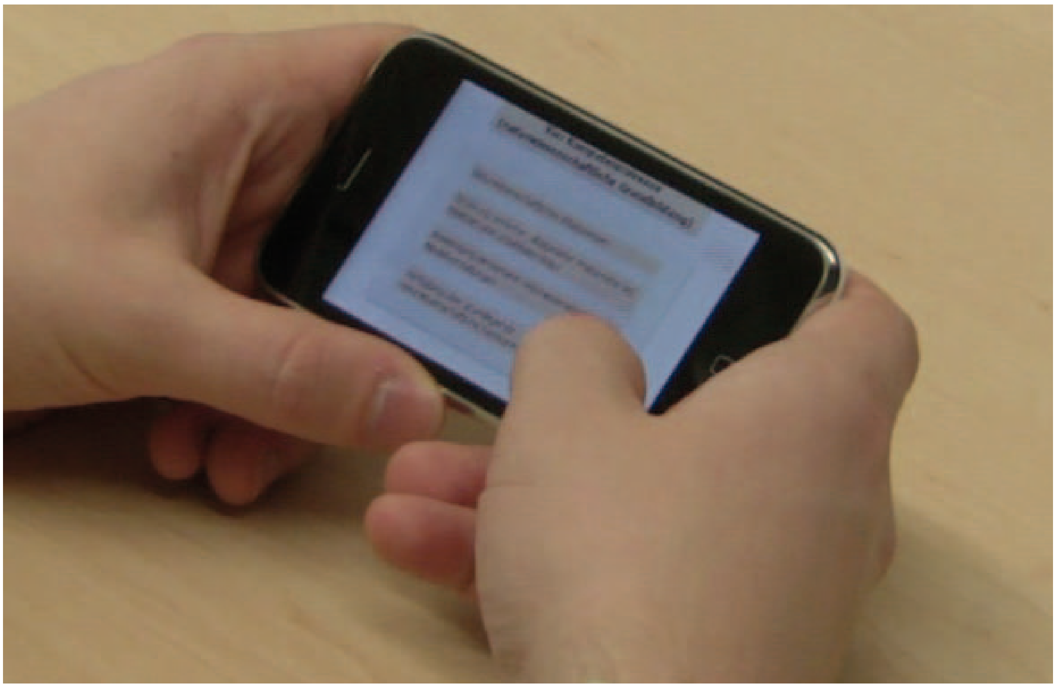


Figure 2.40: Participant holding the iPhone with both of his hands in landscape mode.

interfaces (e.g. the iPhone movie player) do not follow this guideline. Our study shows that users most commonly hold the device in landscape mode for watching movies (cf. Figure 2.40), since this offers the most screen real estate. Moreover, most of them utilized both hands to hold the device. In this case only the thumbs are able to interact with the interface. The rest of the hand is located behind the device. Consequently, the interaction is highly limited by the length of a user's thumb. If a horizontal timeline is used for navigation, such as in the iPhone video player, the timeline should be placed at the bottom of the interface. This allows reaching it with the thumbs and moreover prevents users from occluding the screen while using the timeline for navigation.

2.8 Conclusion

In this chapter, we approached the mobile interaction with large multimedia information spaces from a *device-centric* interaction perspective. For this purpose, we focused on mobile browsing of topically inter-related video collections as a prime example for the mobile interaction with large multimedia information spaces. The aim of this chapter was to adopt a broad view on the design space for mobile video browsing, involving three different classes of navigation complexity: navigating (1) an *individual video segment*, (2) a *larger video* and (3) a *collection of inter-related videos*. This chapter therefore fills a void in previous research, which focused primarily on fine-grained in-scene navigation, and advances the field of mobile video browsing with respect to the following contributions:

- **Exploration of the Design Space for Mobile Video Browsing:** Informed by an analysis of use patterns of mobile video browsing and participatory design sessions, we set up a design space that covers two dimensions: the broad interaction metaphor used in the interaction concept (*GUI-based*, *gesture-based*, *physical*) and the complexity of the navigation.

- **Novel Interface Concepts:** This enabled us to systematically derive 8 interaction concepts (7 novel concepts, one standard interface), which are situated within the design space and implemented on the iPhone. They cover the navigation within individual videos, larger videos and for browsing collections of several inter-related videos.

- **In-depth Evaluation and Detailed Error Analysis:** We conducted a controlled experiment with 44 participants and collected and analyzed more than 18 hours of video observations. Therefore, we were not only able to assess the usability and user experience of each interface, but also to identify where errors occur. The results provide empirical evidence that designers should leverage the novel capabilities of mobile devices, such as direct touch and inertial sensors. A more traditional GUI approach, as in this case the iPhone video player, is likely to lead to lower efficiency and is more error-prone. The usability error analysis shows that even a simple misplacement of interface elements can lead to the loss of internal locus of control and therefore to severe usability breakdowns. Moreover, the error analysis underlines the potential of gesture-based or physical interfaces for mobile video browsing. Our participants committed only little errors of type E1-E3 using either interface type. They mainly committed slips, if at all. These slips can be further reduced when users become more familiar with such novel interaction techniques.

- **Design Principles for Future Mobile Video Browsers:** Our analysis also provided the basis for design principles for mobile video browsers. By supporting spatiotemporal browsing metaphors and discrete temporal navigation and by placing interface elements carefully, designers can improve both usability and user experience of future mobile video browsers.

Future work in this field should consider the further exploration of physical interaction techniques, due to their promising hedonic qualities.

CHAPTER 3

Space-Centric Interaction

In the previous chapter, we have addressed the mobile interaction with large multimedia information spaces from a device-centric interaction perspective. As for the interaction techniques, we particularly constrained the information space to the small display of a mobile device and showed how to leverage the limitations of today's small, mobile devices to foster more usable interfaces and a richer user experience.

In this chapter, we will push the boundaries of the virtual information space toward the physical space and adopt a *space-centric* interaction perspective. For this purpose, we regard the virtual information space as an overlay over the physical space. It is important to note that the virtual information does not necessarily have to be directly bound to a specific physical object. This way, both physical and digital artifacts co-exist equally. Toward this end, we leverage mobile phones as see-through devices for so-called *peephole interaction*. Hence, only a small portion of the large information information space is visible at a time: the part that is overlaid over the part of the real space currently visible in the "peephole" (i.e. the mobile device screen). Users can explore the virtual artifacts through embodied interaction, for instance by panning the mobile phone across the physical space.

This chapter investigates arising challenges in mobile real world settings, such as riding on a bus. Furthermore, we will derive interaction techniques supporting users in these settings. Moreover, we will address an omnipresent issue of embodied peephole interaction: targeting digital artifacts with an a-priori unknown location. We will present an empirically validated mathematical model, which allows us to explain the navigation in one-dimensional information spaces, particularly for a-priori unknown off-screen targets. The model is inspired by physiological aspects of the human body.

This chapter is structured as follows. In Section 3.1, we present an overview over prior work addressing the exploration of digital information in physical space. Furthermore, we point out challenges of embodied dynamic peephole interaction and identify open research questions. Section 3.2 investigates the embodied dynamic peephole interaction in mobile, real world settings. For this purpose, we have conducted a qualitative, exploratory field study. We present the results and derive an empirically grounded theory, characterizing embodied peephole interaction in real world settings. The theory allowed us to derive design implications for future embodied peephole interfaces. As a result of our study, one apparent issue of embodied peephole interaction was the targeting of digital artifacts located off-screen, as mentioned above. Section 3.3 addresses this issue and contributes an empirically grounded mathematical model for the navigation in one-dimensional information spaces, particularly for targets with an a-priori unknown location. The results of a controlled experiment with 32 participants using a physical apparatus validate our model for the navigation in a-priori unknown information spaces. We found that a user's familiarity with the information space and her initial search direction has a significant impact on the navigation time to hidden targets. The chapter concludes with a discussion and an outlook upon future research directions in Section 3.4. In summary, the main contributions of this chapter are

- an empirically grounded theory for embodied peephole interaction in real world settings through a qualitative, exploratory field study,
- design implications for future embodied peephole interfaces,
- a novel model for the embodied navigation of a-priori unknown information spaces with spatially-aware displays and
- results of a controlled experiment with 32 participants using a physical apparatus, validating the mathematical model.

3.1 Overview

In recent years, there has been a larger effort to bridge the physical-digital divide for mobile interaction and mix both physical and digital realities [Billinghurst, Kato, and Poupyrev, 2001, Wagner and Schmalstieg, 2006, Zhou, Duh, and Billinghurst, 2008, Klein and Murray, 2009]. Milgram and Kishino

Figure 3.1: Mixed reality continuum by Milgram and Kishino [1994]

[1994] described the degree of augmentation as the mixed reality continuum. Figure 3.1 illustrates the continuum: the real environment is set at one end of the axis, whereas the virtual environment designates the other end. Augmenting the real environment with digital artifacts, e.g. through projection, then is described as *Augmented Reality* (AR). Augmenting virtual environments with real world cues is described as *Augmented Virtuality* (AV), respectively. To give a simple example for the latter case: when a user is sitting on a chair and immersed in a virtual environment, e.g. through wearing a head-mounted-display, a coupling between actions in the virtual environment and an actuation of the chair resembles an augmented virtuality experience. The boundaries between AR and AV are fuzzy and the space in-between is therefore called *Mixed Reality* (MR) and characterizes the mixed reality continuum.

The benefits of merging the two worlds toward such a mixed reality are obvious: a more immersive user experience and a more "natural" interaction with digital artifacts, which are usually restricted to a two-dimensional display. After giving a broader overview over related work on the exploration of digital information in physical space, we will focus on one specific sub-issue of mixed reality interaction: embodied peephole interaction.

3.1.1 Exploration of Digital Information in Physical Space

The two major challenges when interacting with large virtual information spaces on mobile devices are the exploration and therefore also the navigation. By mapping virtual information to the physical space, physical interaction techniques can be used for this very purpose. Research in this area is diverse: projects such as Sweep-Shake [Robinson, Eslambolchilar, and Jones, 2009] for instance map digital artifacts to geographic locations. With the help of a tangible device, comparable to a dowsing rod, users can then walk around in physical space and explore the digital information. Whenever they are near a digital artifact, the device provides feedback (e.g. vibro-tactile) and thus fosters the user's awareness. Tangible devices have also been explored as navigational aid in TactileBelt [Pielot et al., 2011]. Here, a belt with vibro-tactile feedback is used to guide users from source to destination while traveling shorter distances. Other application scenarios comprise location-based games [Magnusson et al., 2011] or media façade interaction [Boring et al., 2011].

When focusing on visual feedback, two interaction techniques have gained particular attention for interacting with virtual information spaces in physical space: peephole interaction [Yee, 2003, Butz and Krüger, 2003, 2006] and flashlight interaction [Cao and Balakrishnan, 2006]. Both are based on a similar idea. They assume the mobile device to be situated in physical space (as a so-called spatially-aware display [Fitzmaurice, 1993]). The device can then be moved in space as a see-through device to explore a virtual information space in the real world. In case of peephole interaction, the screen of the mobile device is used as a window that overlays the physical space with virtual information. The flashlight metaphor utilizes mobile projectors to display the information space. However, both techniques reveal only a part of the virtual information space to the user. Both approaches have their advantages and disadvantages. While a projector-based approach allows for a more immersive coupling between virtual and physical space, the projection is publicly visible. Correspondingly, using mobile devices as peepholes allows for a more private interaction but also limits a user's view upon the virtual space to the mobile device. In the following section, we discuss research related to peephole interaction. Projector-based interaction is discussed in detail in Chapter 4.

3.1.2 Embodied Peephole Interaction

The original idea of using mobile devices as peepholes and information lenses dates back to Fitzmaurice's seminal work on spatially-aware displays [Fitzmaurice, 1993]. Fitzmaurice mainly focused on enriching physical objects such as book shelves with additional, context-dependent information, e.g. for books located in the shelves. Conceptually, there are two different modes of interaction: *static* and *dynamic* peephole interaction [Mehra, Werkhoven, and Worring, 2006]. On the one hand, static peephole interaction assumes that the

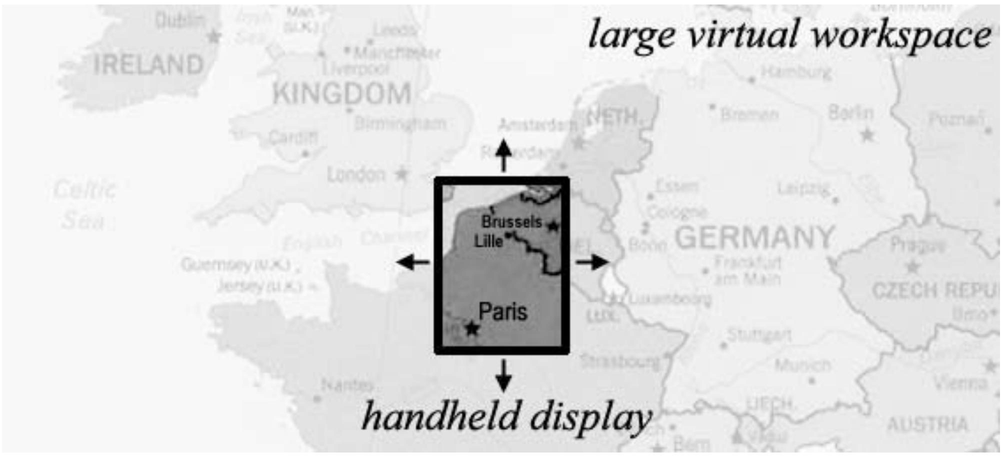


Figure 3.2: Peephole concept: a small handheld display is held over a larger virtual information space which allows for spatial exploration [Yee, 2003].

information space is moved, not the peephole (comparable to moving the map in Google Maps). On the other hand, dynamic peephole interaction focuses on moving the peephole, i.e. the mobile device.

Information Lenses. The concept of information lenses was in consequence refined by Rohs et al. [Rohs and Essl, 2006, 2007, Rohs and Oulasvirta, 2008, Rohs et al., 2009]. They for instance investigated how additional information can be displayed with respect to underlying physical paper maps. In particular, a user can intuitively understand the limits of the virtual information space, due to its strong coupling to the paper map. Using mobile devices as lenses has also been explored in [Looser et al., 2004, Schmalstieg et al., 2002] and also found its way toward commercially available products such as Layar.com and Wikitude.com for geo-location-based additional information.

Peephole Interaction. Yee [2003] investigated various interaction techniques for peephole interaction. Yee's work acknowledged the fact that the virtual information does not necessarily have to be directly bound to a specific object, thus resemble only additional information. He particularly focused on designing techniques, which allow users to explore a personal information space surrounding them, but which is not directly bound to e.g. their office desk. Moreover, he considered the virtual information space as potentially unlimited.

Embodied interaction with peepholes was also explored by Luyten and colleagues who investigated the use of dynamic peephole interaction for computer supported cooperative work [Luyten et al., 2007]. They addressed the problem of using multiple peepholes while interacting within the same virtual information space. Furthermore, they proposed methods to solve concurrent manipulation problems, such as cooperatively interacting with the same digital artifact.

Peephole interaction has most recently been explored for tangible interaction by Spindler, Tominski, Schumann, and Dachselt [2010]. They built so-called tangible views, allowing them to explore the space above a tabletop system and provide support for tangible focus+context or overview+detail interfaces [Cockburn, Karlson, and Bederson, 2008] on interactive surfaces.

Summary. Since only a limited part of the information space is visible through the small window, loss of orientation is clearly an issue [Jul and Furnas, 1998, Edwards and Hardman, 1999]. While visual cues for information navigation have been explored in Halo [Baudisch and Rosenholtz, 2003] or Wedges [Gustafson et al., 2008] (see also the comparison by Henze and Boll [2010] and [Cockburn, Karlson, and Bederson, 2008] for a holistic overview over visual cueing techniques), optimizing the layout of the information space is essential for increasing the efficiency of a user. The layout can be optimized by considering navigation times to targets depending e.g. on their distance. However, it is still unclear how users actually perform when navigating toward targets that they cannot see upfront.

In summary, there have been various approaches which utilize mobile devices as a spatially-aware displays for peephole interaction. As a matter of fact,

- These approaches mainly focused on using the mobile devices as information lenses. This generates a strong coupling between physical and digital artifacts. Yee's work acknowledged the fact that the virtual information does not necessarily have to be directly bound to a specific object, thus allowing (1) for virtually unlimited information spaces and (2) physical and digital artifacts to co-exist equally. He concentrated on the design of fundamental interaction techniques.

- These fundamental techniques were evaluated in lab settings and not in a real world context, where they actually should be applied. This is particularly important since the information space is dynamically mapped to the physical space and its actual representation strongly depends on the highly dynamic situation.

- Last, previous research focused on visual cueing techniques to scaffold the exploration of information spaces and neglected that optimizing the layout of the information space is equally essential for increasing the efficiency of a user. The layout can be optimized by considering navigation times to targets depending e.g. on their distance. However, it is still unclear how users actually perform when using embodied peephole interaction for the exploration of large information spaces.

The community hence lacks a fundamental understanding of how users would actually interact with such an immersive representation of the information space. We particularly investigate mobile settings and therefore focus on *dynamic* peepholes. These observations lead us to the following research questions:

- How do users interact with an embodied dynamic peephole in mobile, real world settings and how do they cope with highly dynamic situations, e.g. in public?

- How can the navigation using embodied dynamic peepholes be modelled to better understand a user's performance?

We have addressed the first question in a qualitative, exploratory field study. The results, an empirically grounded theory and implications for interaction techniques, are presented in Section 3.2. The second question is investigated in Section 3.3, where we contribute a mathematical model for one-dimensional embodied peephole pointing, where the location of the targets is a-priori unknown. The model has been verified and compared to related approaches empirically in a controlled experiment.

3.2 Exploratory Field Study

We conducted an exploratory field study to investigate the embodied dynamic peephole interaction in mobile, real world settings. The study investigates potential problems that users might encounter in different real world scenarios; e.g. contrasting experiences from closed environments such as office spaces with rather noisy environments such as public places. In the vein of Yee's work, we focused on dynamic peephole interaction with potentially unlimited virtual information spaces, not bound to a specific physical object. Hence both physical and digital artifacts co-exist equally.

Attributing to the three-dimensional nature of the physical space, we chose three-dimensional knowledge networks as virtual information spaces. The choice for knowledge networks was also motivated by their simplicity in visualization, as well as being a particularly pertinent example of large, complex information spaces. Figure 3.3 shows an example of such a knowledge network.

Figure 3.3: Example visualization of a large, graph-like information space which is mapped to the physical space. In this case, the graph resides on an office desk.

3.2.1 Study Design

We recruited 12 participants (10 male, 2 female) between 20 and 60 years of age (mean 28 years). Half of the participants were left-handed. All of the participants were knowledge workers with different scientific backgrounds (i.a. physics, civil engineering, computer science, pedagogy and politics). We chose a within-subject design. The participants had to explore various information spaces in different settings: (1) in their workplace, (2) on the train and (3) outdoors. We selected these places due to their spatial and social framing. We were particularly interested in the differences between a physically constrained and dynamic situation such as on a train (with strangers being present and only little space for interaction) and a more open and static situation, such as in a workplace or even outdoors.

As information spaces, we chose low-fidelity prototypes of three-dimensional knowledge networks. The graph-like structure represented the table of contents of the Wikipedia article on “Social Media” [1]. Hence, each level of the graph contained the entries in the corresponding level of the table of contents. The knowledge networks were only mock-ups and therefore non-interactive. We opted for a low-fidelity prototype to encourage the participants’ imagination. Moreover, we did not want to influence the participants by any design, since we wanted to explore fundamental dimensions for interaction and visualization. The mock-ups were displayed on an iPhone, which the participants used for the tasks (cf. Figure 3.4). The mock-up always showed an excerpt of the information space: the actual view of the peephole. Thus in particular, it did not show the whole information space. The concept of embodied peephole interaction was explained to them upfront. Moreover, to give the participants a basic impression of how the user interface is adapted according to the embodied interaction with the device, they were presented with video prototypes of the interface animation.

The order of both settings and data sets was fully counter-balanced. Each single-user session lasted about 1.5 hours.

[1] http://de.wikipedia.org/wiki/Social_Media (last checked: October 29, 2012).

Figure 3.4: Participant interacting with an interface mock-up displayed on the iPhone. Here, the participant was on a train.

3.2.2 Methodology: Data Gathering and Analysis

We chose a qualitative data gathering and analysis methodology [Strauss and Corbin, 2008]. The method was inspired by Grounded Theory [Strauss and Corbin, 1990] and performed iteratively per session. As data gathering methodologies, we used semi-structured interviews, observation and photo documentation. Moreover, the participants were asked to think aloud. However, the main object was to observe the participants while interacting with the iPhone, where the interface mock-ups were visualized. During the interviews, the participants were given the possibility to sketch their ideas as UI paper prototypes.

In addition to observing the general interaction, we wanted to compare and contrast different visualizations containing different visual cues for off-screen elements. In each setting, the participants were presented with four different visualizations as shown in Figure 3.5. The first two visualizations (cf. Fig 3.5 a and b) show an undirected and directed hypergraph, respectively. The directed hypergraph contains directed arrows as edges. These give a directional hint with respect to two adjacent vertices. Moreover, it can be interpreted as a visual cue, indicating the location of off-screen elements. The third visualization (cf.

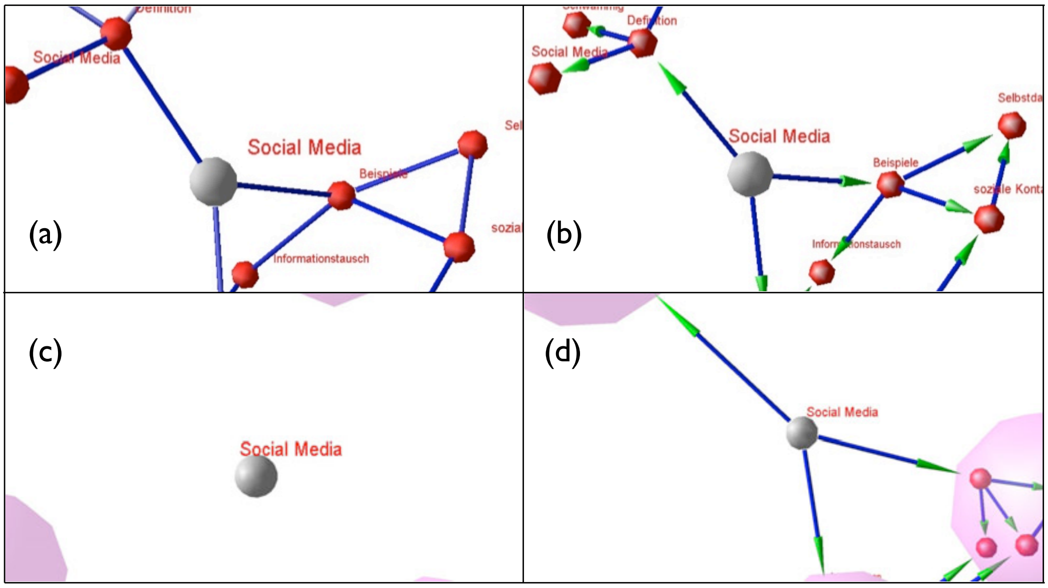


Figure 3.5: Overview over the mock-up visualizations: (a) undirected hypergraph, (b) directed hypergraph, (c) a graph containing only vertices, halos indicating further content and (d) directed hypergraph with halos indicating further content.

Fig 3.5 c) is essentially a graph containing only vertices, with three-dimensional halos (in the sense of Baudisch and Rosenholtz [2003]), indicating off-screen vertices. The fourth visualization (cf. Fig 3.5 d) is a directed hypergraph with 3D halos.

The participants were asked to imagine that when holding the device right in front of their upper body, they could see the current view of the peephole display. As a next step, they were asked to navigate to another, specific vertex in the knowledge network so that it would reside in the center of the peephole eventually. As mentioned earlier, the knowledge network depicted the table of contents of the Wikipedia article on “Social Media”. To give a simple example: the root vertex was designated as the first level of the table of contents, hence labeled “Social Media”. The participants were then for instance asked to move to the “Definition” of social media–an entry on the second level of the table of contents (see Fig. 3.5.b, “Social Media” in the center of the screen, “Definition” is located in the top-left corner).

Of course, the mock-up did not follow the embodied movements of the participants. However, the main objective here was to see how they would intentionally perform the interaction (e.g. by solely moving the hand/arm or even walking toward the vertex). In the workplace and on the train, the participants were asked to perform the navigation tasks while sitting, as well as standing.

After each session, the interviews and observations were transcribed. Salient quotes were selected and analyzed using an open, axial and selective coding approach [Strauss and Corbin, 2008]. The emerging categories served as direct input for the follow-up session with the next participant. The scope of the session was adapted according to the theoretical saturation of the categories.

3.2.3 Results and Discussion

The analysis and coding process yielded four major categories, characterizing the embodied peephole interaction with knowledge networks in real world settings: *contextual constraints*, *user preferences*, *visualization* and *navigation*. In the following, we will report our findings within these categories. Moreover, these categories also characterize design implications for future interaction designs, presented in the following section.

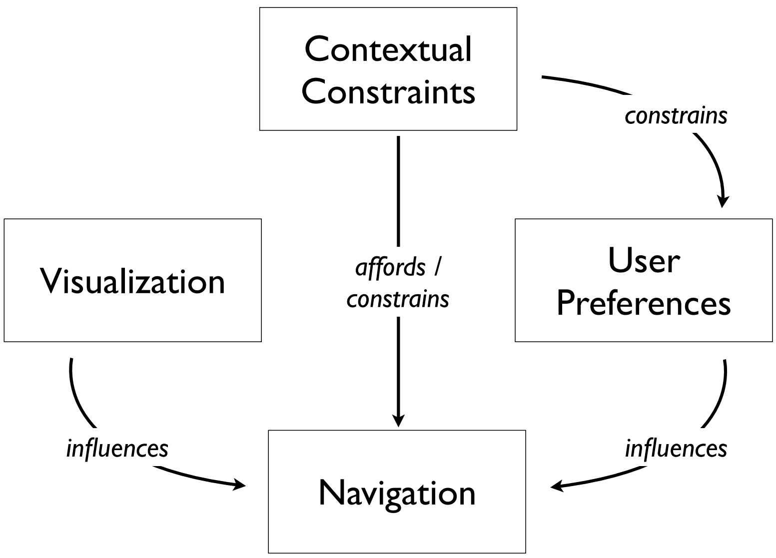


Figure 3.6: Categories and their interrelationship.

3.2.3.1 Contextual Constraints

Throughout the study, we noticed that contextual constraints have a large impact on how users interact within a mobile, real world setting, and affect both navigation and user preferences. We subdivided this category into two different types of contextual constraints: *physical* and *situational* constraints.

With respect to physical constraints, we found that the *physical space* provided for interaction influenced how the participants interacted with the embodied peephole. Figure 3.7 shows a participant sitting in a train. The participant is right handed and here, he held the device in landscape mode. In this very situation, the target vertex was located to his left hand side. Our observations revealed two phenomena:

1. The participant imagined that the vertices to his left side would collide with the train's glass window. He further suggested that in this very situation, he would appreciate to "move" the whole information space to the right, since he wanted to omit the collision with the window.
2. He further imagined that the vertices would then be displayed on the windows' surface, since if they had been displayed further left, they would had been outside the train. He thus bent his left shoulder a bit back and then explained that he would take a look at the vertices like he did in Figure 3.7.

Situational constraints were also apparent. We observed that when the participants were able to *walk* within the information space, e.g. in the outdoors setting, the interaction was completely different compared to the interaction on the train or in the office. We observed that once the participants were standing and had enough physical space (e.g. outdoors), they utilized the movement of the upper body to position the mobile device (*target acquisition*) and afterwards, they walked toward the target (*target selection*). One participant commented that "*when walking, the diameter of the information space can be nearly unlimited and the space can be explored rather freely, as opposed to manual interaction with one or two hands, where the interaction space is limited by one s arm length*". We observed this also in the office setting, where participants were first seated and then stood up to walk toward a target. However, one participant noted that he needs "*a good reason to not just stay seated,*

Figure 3.7: Participant navigates to a vertex in the knowledge network to his left hand side. He information space collides with the physical space, here a train's glass window. He therefore bends his shoulder a bit to the back and imagines the information space to be displayed on top of the window's surface.

since it is far more comfortable to not have to walk toward a target, but to just grab it or even have it near oneself". While this preference is rather obvious, it shows that designers need to adapt their interfaces to user preferences and pay careful attention to contextual constrains.

3.2.3.2 User Preferences

As user preferences, we mainly observed how the participants held their device, i.e. both the device orientation and the user's handedness. This is constrained by the context, e.g. the participant in Figure 3.7 was unable to utilize his left hand for the interaction, although he is left-handed.

We observed that the *device orientation* influenced how the participants actually moved the device through physical space. Interestingly, the device was always moved vertically to explore the information space when it was held in landscape mode. When it was held in portrait mode, the respective participant started to explore the information space in the horizontal dimension.

Moreover, we found that the *user's handedness* correlates to the preferred direction for the exploration of the interaction space. As one participant noted: "*Since I am right-handed, I do not want to move the device to the right hand side. My upper body movement is constraint to the right. So I tend to move the device to the left-hand side, because I can move it more freely there*". We observed the corresponding effect also for the left-handed participants. They preferably started to explore the information space to the right hand side, since their upper body blocked the movement to the left hand side to some extent.

3.2.3.3 Visualization

In the interviews, we found that the visualization influences the way users actually perceive the physically embedded information space and in turn influences the navigation. Poorly positioned content might be unreachable or directional cues might be misinterpreted. Key sub-dimensions here are *visual layout* and *level of abstraction*.

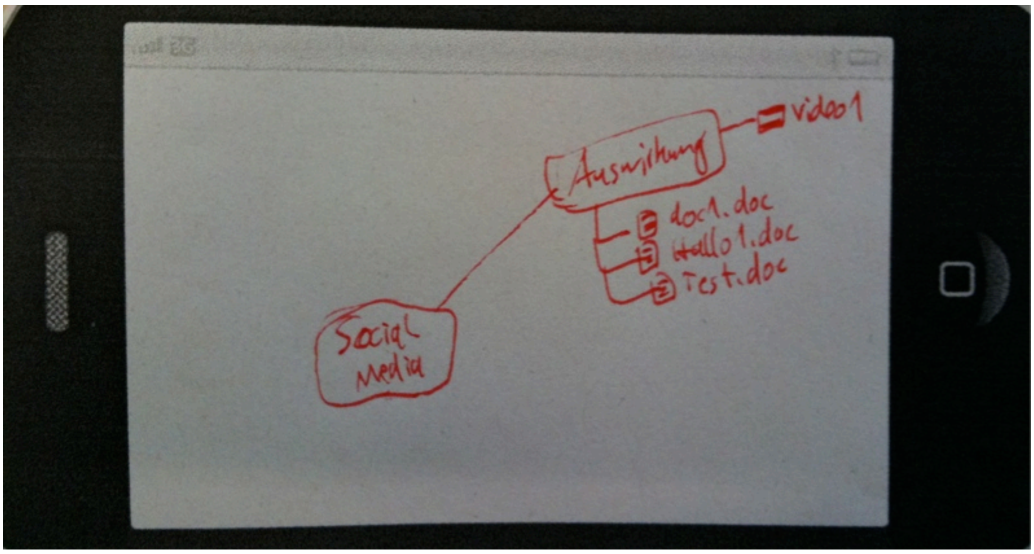


Figure 3.8: User interface sketch by one of the participants. The sketch illustrates different levels of abstraction for the knowledge network. Media, such as text documents or videos are located in the leaf nodes, which can be expanded and collapsed by tapping onto their respective root vertex.

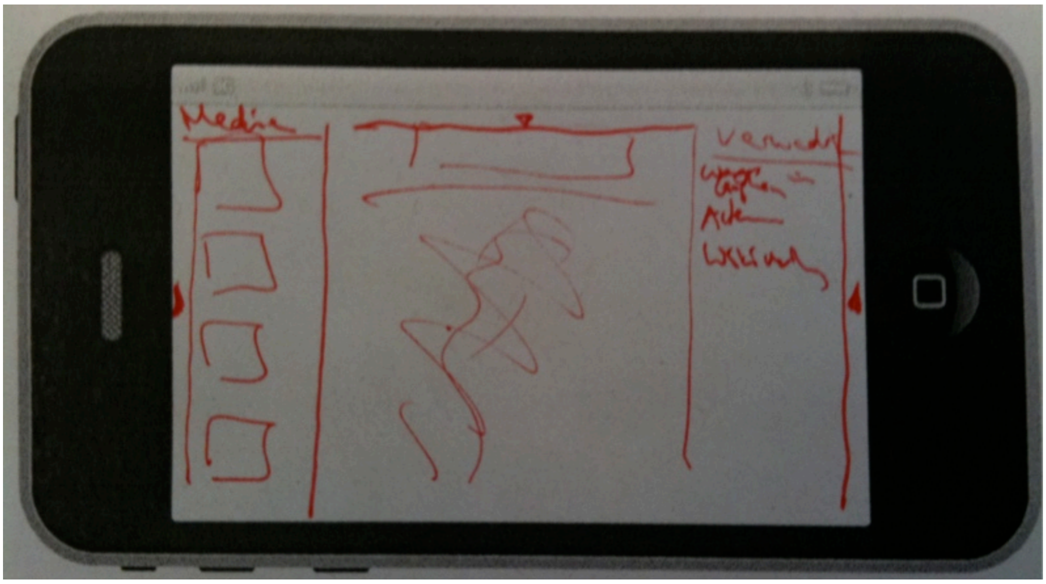


Figure 3.9: User interface sketch by one of the participants. The participant added two-dimensional side-bars to the peephole visualization (here: in the center). The left sidebar displays further content with respect to the currently focused vertex. Related vertices, e.g. those which are adjacent, are additionally displayed in the right side-bar to provide quick access.

The participants reported that the directed graph visualization supports their orientation. This was mainly due to the directional hints implied by the directed edges. One participant noted that "*the directed edges help to estimate the target location in three-dimensional space*". The participants also appreciated the halo visualization in combination with the directed graph: "*this groups the vertices, which provides a better overview for me*", as stated by one participant. However, the visualization relying solely on halos was perceived as inferior. The participants explained that this particular visualization does not support the three-dimensional navigation. A few participants described the user experience as "*feeling trapped*", since the halos "*take a major part of the interface and constrain the viewport*". The undirected graph was also perceived as inferior to the directed graph due to not providing directional cues through the edges.

Regarding the level of abstraction, one participant suggested to use collapsable vertices in the knowledge network. Figure 3.8 shows a sketch of his idea. He imagined the vertices to contain for instance multimedia content such as text documents or videos. And for instance by tapping onto one specific vertex, children of that very vertex can be interactively expanded (here: "Video1, Doc1.doc, Hello1.doc and Test.doc") or collapsed, respectively.

Another participant suggested to provide contextual information while navigating physically through the information space. Figure 3.9 shows his user interface sketch. He imagined to have two additional sidebars on the left and the right hand side of the interface. The knowledge network is to be displayed in the center of the interface. The sidebars visualize additional information with respect to the currently focused vertex. For instance, additional multimedia information can be displayed on the left hand side. A user can then e.g. play back a video by tapping onto the respective icon. The participant further imagined that related content (i.e. adjacent vertices) can be visualized on the right hand side and by tapping onto one related vertex, that very vertex can be shifted into the center of the interface.

3.2.3.4 Navigation

The aforementioned dimensions influence the way the participants imagined to navigate the virtual information spaces. We observed that the participants imagined the information space to be situated in physical space in basically two different modes: either *user-centric* or *space-centric*.

In case of a user-centric point of view, the participants assumed that the space is centered around themselves (and also moves with them) like for instance a virtual bubble. Thus, they imagined themselves to be the actual reference point. This is comparable to Pederson and Surie's "egocentric perspective" [Pederson and Surie, 2007] on the interaction with everyday objects, where the user as well serves as the centre of reference to all of the user's interactions.

The space-centric view, implied that the centre of the information space is anchored in physical space. For instance one participant imaged the centre of his room to serve as the center of the virtual information space. He then could walk through the information space by walking through his room.

How the participants perceived the mapping of the virtual to the physical space was mainly determined by the contextual constraints. For instance on the train, the participants mainly adopted an ego-centric perspective, whereas they adopted a space-centric view outdoors, where they could walk. The perception was indifferent in the office setting.

Our observations further revealed that the navigation in the z-direction (i.e. zooming) also depends on the contextual constraints. In case the participants were standing, they preferred to move the device *toward their target*. In case they were seated, they assumed that moving the peephole *toward their eyes* designates zooming. One participant stated that it is comparable to "*reading a paper versus reading a sign post, the sign post is xed and I need to walk toward it, whereas I can just take the piece of paper and lift it toward my head and read it*".

3.2.4 Implications and Conclusion

The previous section described four relevant theoretical dimensions when designing interactions for embodied dynamic peephole interaction. The actual navigation is constrained by physical, as well as situational constrains and influenced by both visualization and user preferences. The inter-relationships of these categories allow us to derive design implications.

3.2.4.1 Exemplary Interaction Techniques

In the following, we showcase how each of the categories depicted in Figure 3.6 influence potential interface designs, by describing two interaction techniques which we implemented for the iPhone. The first, *Acquire and Zoom*, exemplifies how user preferences and visualizations impact the navigation (cf. Figure 3.10). The second, *Grab and Rotate*, shows how contextual constrains can be overcome to scaffold a user's navigation (cf. Figure 3.11).

Acquire and Zoom Whether users are seated or standing influences the way they want to navigate the information space (***User Preference***). When e.g. standing, users utilize the movement of the upper body to position the mobile device (*target acquisition*) and afterwards, they walk toward the target (*target selection*). In contrast to this, walking is not an option to navigate the information space when users are seated. Hence, they need to be able to acquire certain sub-areas of the information space and then zoom toward them while seated to continue browsing the information space. To support this, we suggest to use direct touch gestures followed by embodied interaction: first, the user

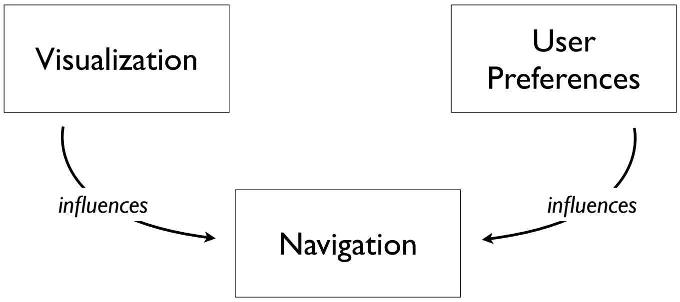


Figure 3.10: Relevant categories for Acquire and Zoom.

acquires the sub-area of the information space and points the peephole toward it. By touching onto the display and dragging the whole device toward the user, the interface zooms into that very sub-area (***Visualization and Navigation***).

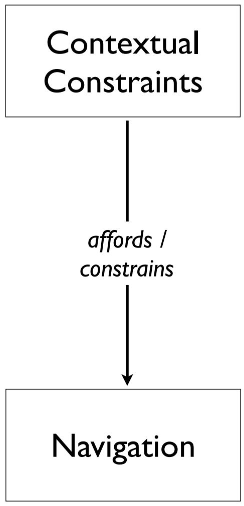


Figure 3.11: Relevant categories for Grab and Rotate.

Grab and Rotate The surroundings and particularly the provided interaction space have a huge impact on how freely users are able to interact with the information space. For instance on the train, the interaction space is rather limited (***Contextual Constraints***). To overcome the constraints in spatial navigation, users need to be able to reposition (e.g. the rotation or simply the movement) the information space dynamically (in analogy to the mouse movement on a mouse pad, when the mouse reaches the border of the mouse pad). This is also a technique which can be applied to overcome collisions with the physical world (as it was e.g. the case in Fig. 3.7). Figure 3.12 illustrates our exemplary interaction technique. The user first targets a sub-area of the information space and touches onto the display (1). Using this gesture, the user grabs and locks the information space in hand. Moving the device (2+3) now repositions the whole information space. By releasing the thumb (4), the information space is placed in-situ (***Navigation***).

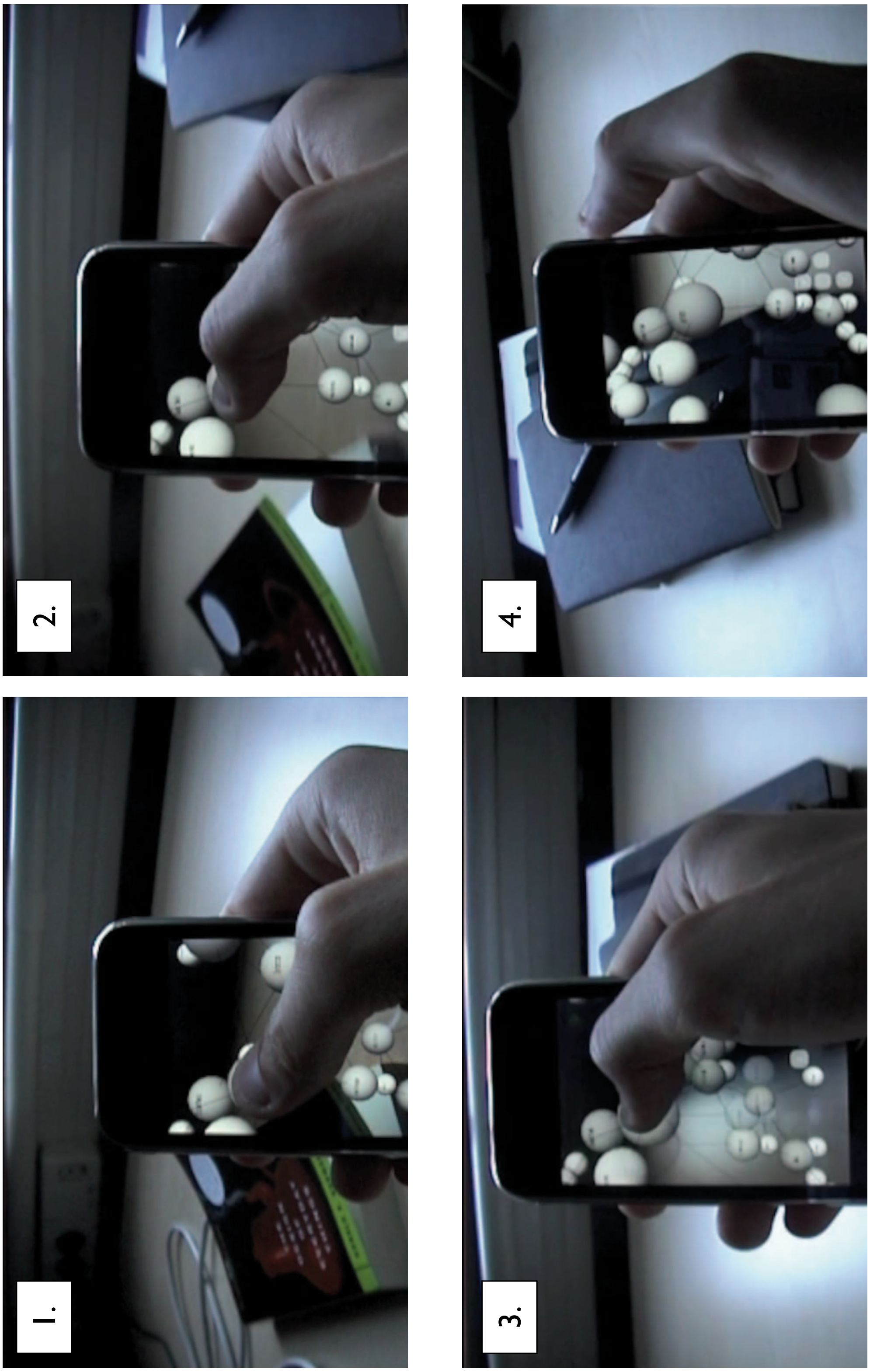


Figure 3.12: Grab and Rotate interaction technique for repositioning the virtual information space in physical space.

3.2.4.2 Summary

The results from our study are summarized in Table 3.1. They underline that designers need to carefully consider contextual constraints, user preferences and the type of visualization when designing interfaces for embodied peephole interaction. For instance, we compared different visual cues to scaffold the navigation: undirected and directed edges, as well as Halos. The results

Categories	Description
Contextual Constraints	
˪ Physical	The provided physical space may restrict the interaction and can collide with the virtual space.
˪ Situational	Standing affords walking for embodied navigation. Sitting requires alternative techniques such as *Acquire and Zoom*.
User Preferences	
˪ Device Orientation	Landscape mode affords vertical embodied panning; portrait mode affords horizontal embodied panning.
˪ Handedness	Restriction of upper body movement due to right- or left-handedness determines initial navigation direction.
Visualization	
˪ Visual Layout	Directional cues scaffold the embodied navigation in three-dimensional space.
˪ Level of Abstraction	Contextual 2D information can interactively enrich the navigation experience and its usability.
Navigation	
˪ User-Centric	Information space evolves around the user and moves with her. The user is the center of the reference.
˪ Space-Centric	Information space is anchored in the physical space, e.g. in the center of a room.
˪ z-Direction	Navigation in the z-direction (i.e. zooming) depends on situational constraints. When standing, users move the device toward the target. Being seated inverses the direction and the device is moved toward one's eyes.

Table 3.1: Summary of the results and sub-categories with respect to the main categories, as depicted in Figure 3.6.

show that directional cues such as directed edges support the navigation to off-screen elements best. We also found that they particularly supported the three-dimensional spatial navigation which is crucial for embodied peephole interaction in large information spaces.

Findings in the navigation category showed that there are two prevailing models how users perceive an information space: either user-centric or space-centric. The actual strengths and weaknesses of both approaches remain to be investigated. While we focused on the *exploration* of existing information spaces, it is highly interesting to investigate how these two models are applied in day-to-day usage, e.g. when creating or restructuring information spaces through embodied peephole interaction. We speculate that for instance a space-centric model might help in *retrieving* information stored in a particular place in physical space, e.g. when a set of papers relevant to a current work project is always stored near a filing cabinet. Thus, the filing cabinet can serve as a meaningful physical hot spot, allowing for easy and quick access to the virtual documents.

Moreover, it is worthwhile to investigate the effects of dynamic settings on the navigation dimension, e.g. how interactions need to be designed when users often switch from interaction spaces with a high degree of freedom (e.g. when walking) to ones with a lesser degree of freedom (e.g. when seated) and vice versa.

The results also revealed that the layout of the information space in physical space is highly important. Poorly positioned content might not be reachable. Besides addressing this with visual cues [Cockburn et al., 2008], optimizing the layout of the information space is essential to increasing the efficiency of a user. The layout can be optimized by considering navigation times to targets depending i.a. on their distance.

In the following section, we show how navigation times can be modeled for embodied dynamic peephole pointing for a-priori unknown targets, i.e. located off-screen.

3.3 A Model of Embodied Dynamic Peephole Pointing

In the previous section, we investigated embodied dynamic peephole interaction in real world settings. The results showed that understanding a user's preferences and context is crucial to solving layout problems and prevent user's from getting lost in hyperspace. This section focuses on modeling a user's navigation performance for embodied dynamic peephole interaction. We model the performance as navigation times in information spaces with a-priori unknown target locations. Since users navigate through pointing to a target, the problem can be seen as a pointing task. Pointing tasks have been most commonly modeled using Fitts' law. However, in his experiments, Fitts' measured the pointing time between two visible targets, hence an *aimed* movement. In our case, the target is not visible a-priori. Previous research proposed several models for peephole pointing [Rohs and Oulasvirta, 2008, Cao et al., 2008]. However, in all experiments, the participants were given directional or visual hints for the target's location. As a consequence, the empirical data showed a high correlation with variations of Fitts' formula.

We contribute a novel, nonlinear model for the embodied navigation of a-priori unknown information spaces with spatially-aware displays. The exploration of unknown information spaces is relevant for many novel location-aware applications such as handheld augmented reality browsers, where users are constantly confronted with new information spaces. A "full-fledged" model of this kind can be quite elaborate, taking into account 2D panning, zooming along the z-axis (depth), shortcut operation, complementing navigation with targets that "approach" the peephole (e.g., bound to device tilt operations), and more. Since such an elaborate model is beyond the scope of this (much broader) thesis, we deliberately constraint our efforts to setting the stage with an initial one-dimensional approach. The model is inspired by physiological aspects of the human body. We conducted a controlled experiment with 32 participants which provides empirical evidence that this type of navigation does not follow a Fitts' law–as claimed in previous experiments.

The remainder of this section is structured as follows. We first discuss related modeling approaches. Next, we describe our theoretical model and illustrate its relevancy for embodied interaction. We then present the results of our experiments. We conclude the section with a discussion of the results and an outlook upon future work.

3.3.1 Related Approaches

There is a large body of knowledge for movement time models. Every model is heavily influenced by the employed input technique and physiological aspects like vision and the human motor system. There exist models e.g. for pointing [Cao et al., 2008, Rohs and Oulasvirta, 2008], scrolling [Andersen, 2005] and aimed movement [Fitts, 1992].

Fitts' original work [Fitts, 1992] focused on one-dimensional aimed movement, where subjects were asked to tap two visible targets consecutively. His model predicts the movement time T to a target of width W in dependency of the target distance D. The model he derived is typically formulated as $T = a + b \log_2(1 + D/W)$, where the logarithmic term defines the index of difficulty (ID) and both a and b are empirically determined constants. There exist various different interpretations of Fitts' law [Drewes, 2010] and it has also been extended to higher dimensions [MacKenzie and Buxton, 1992]. This model of aimed movement inspired movement time models for input techniques such as peephole interaction, e.g. by Cao et al. [2008] or Rohs and Oulasvirta [2008].

Cao et al. [2008] conclude that one-dimensional dynamic peephole pointing follows a Fitts' law with

$$T = a + b \left(n \log_2(\frac{D}{S} + 1) + (1 - n) \log_2(\frac{D}{W} + 1) \right), \tag{3.1}$$

where S designates the window size and n is empirically determined. In their experiment, they simulated peephole pointing tasks using a 22" screen as their information space. The peephole as a window onto the information space was controlled using a stylus and a Wacom tablet. Targets were simple vertical lines of a certain width and "infinite" height. Although the targets in the experiment were truly hidden, the participants were given a directional hint towards the target location. However, the experimental setup does not consider embodied interaction with a physical display (which in turn has a certain friction and acceleration). Moreover, the Wacom tablet only provides a small interaction space which does not allow to cover physical constraints imposed by human physiology, which are crucial for embodied interaction.

The work by Rohs and Oulasvirta [2008] targets embodied interaction using mobile devices in two-dimensional space (see also [Rohs et al., 2011]). They claim that dynamic peephole pointing clearly follows a variation of Fitts' law (also with a logarithmic ID), particularly for the case when the targets are not visible to the users. However, in their experiment, the participants were a-priori aware of the target's position and their actual task was rather an aimed movement, which consequently implies a high correlation with Fitts' formula.

Andersen [2005] derived a movement time model for one-dimensional scrolling tasks. Scrolling tasks are quite similar to peephole pointing tasks, since users only see a small portion of the information space (e.g. a browser window). Andersen's experiment contained implicit hints by letting participants either begin at the top or at the bottom of a document. Andersen found that scrolling for hidden targets does not follow a Fitts' law, since (1) the targets are obviously not visible as opposed to Fitts' experimental setup and (2) the actual movement time is limited by human perception, namely the maximum rate at which a target can be perceived when scrolling by. Andersen found that 1D scrolling tasks follow a simple linear model:

$$T = a + b\,D, \tag{3.2}$$

with D being the distance to the target and a and b being empirically determined constants.

3.3.2 Theoretical Model

In the following, we derive a model for one-dimensional dynamic peephole pointing with spatially aware displays where the location of the target is a-priori unknown. Users employ embodied interaction to navigate the information space, therefore moving the whole display (or device) through space. The latter fact implies that the navigation within the information space is highly depending on the human motor system. Users hold the device in hand and move it horizontally along an axis to browse the one dimensional information space (see Figure 3.13). Without loss of generality, we assume that the width of the information space is limited by one's arm length and that one starts exploring

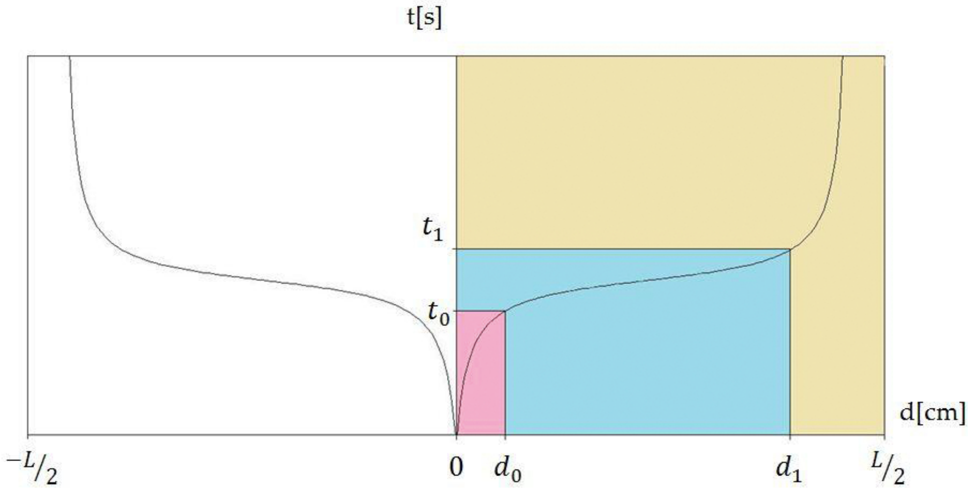


Figure 3.13: Model for the arm movement of a spatially aware display in one-dimensional space. The vertical axis denotes the time required for the movement to a position. The display is moved along the horizontal axis.

it in the middle. Hence, the user has the possibility to either start navigating towards the right or the left hand side. We moreover assume that the user starts her navigation in the center of the information space.

We claim that the device movement time using embodied interaction is trigonometrically dependent due to (1) a non-linear acceleration when moving ones arm and a slightly slower movement directly in front of ones upper body (see interval 0 to d_0 in Fig. 3.13), (2) a more flexible movement when the arm is near one's upper body (hence the arm is relaxed and not sprawled, see interval d_0 to d_1 in Fig. 3.13) and (3) a decrease in movement speed after the arm has been sprawled out to a certain degree (see interval d_1 to $\frac{L}{2}$ in Fig. 3.13. This movement represents a tangent, shifted by half a period, scaled to match the interval of $(-\frac{L}{2}, \frac{L}{2})$ and depicts the index of difficulty (ID) in the sense of Fitts (cf. equation 3.3). $L \in \mathbb{R}$ is the width of the information space. The target's distance is $D \in \mathbb{R}$ and naturally limited by L.

$$ID = \tan\left(\frac{D}{L}\pi + \frac{\pi}{2}\right). \tag{3.3}$$

The display is of width $S \in \mathbb{R}$. This leads to the following formulae for the movement time $T \in \mathbb{R}$, with $a, b \in \mathbb{R}$ being empirically determined constants:

$$\begin{aligned} T &= a + \frac{b}{S} ID \\ &= a + \frac{b}{S} \tan\left(\frac{D}{L}\pi + \frac{\pi}{2}\right). \end{aligned} \tag{3.4}$$

3.3.3 Experiment

Prior research on the movement analysis of dynamic peephole pointing has focused either on tasks with prior knowledge of the target location and thus an aimed movement or on other specific interaction techniques, not on embodied interaction. Moreover, directional cues regarding the target location were given in all studies. We therefore wanted to particularly investigate the impact of the human physiology on embodied dynamic peephole pointing and the case without any directional cue. Thus, we verified the following hypotheses with a controlled experiment:

H1: The character of the information space (familiar versus unfamiliar contents) affects the search time.

H2: Embodied dynamic peephole pointing for hidden targets in an unfamiliar information space is neither sufficiently modeled by Fitts' law or its derivates (e.g. Cao et al. [2008]), nor by a linear model.

H3: A larger target distance results in a larger search time.

H4: A larger peephole results in shorter search times.

H5: Starting in the wrong direction will add a constant factor to the time taken when starting in the correct direction.

3.3.4 Experiment Setup and Methodology

Apparatus. We have designed a physical apparatus (see Figure 3.14) for our experiment instead of utilizing a handheld device. A physical apparatus allows for embodied interaction and ensures a reliable in-/output with minimal noise (which can be easily caused by hand jitter when using a handheld device in 3D or due to tracking errors). We thus are able to abstract from a concrete interaction metaphor such as a flashlight interaction with a mobile projector, without losing general applicability of our results for embodied dynamic peephole interaction.

The apparatus consists of a $1,40\,m$ long and $10\,cm$ wide rail and a belt with an exchangeable plastic window. The participants were apparently able to estimate the length of the rail, but this does not impact our experiment, since we assume that the length of the information space is limited by one's arm length. The window was used as a peephole onto a strip of paper, representing the information space. The targets were printed onto the strip (see Fig. 3.15). The window was equipped with a handle, enabling the participants to slide the window along the rail, thereby revealing the targets. The physical apparatus was designed such that the acceleration is comparable to that of a handheld device used for embodied interaction. Moreover, once accelerated, there was virtually no friction, comparable to the movement of a handheld device in the air.

Figure 3.14: Apparatus: a $1,40$ m long and 10 cm wide rail and a belt with an exchangeable plastic window.

Figure 3.15: Participant manipulating the window. The targets were printed onto a physical paper strip.

Participants. We have conducted a controlled experiment with 32 participants (10 female, 22 male, 29 right- and 2 left-handed). The window handle was positioned according to the handedness (i.e. left for a left-handed participant) to exclude any effects through e.g. occlusion. The age of the participants ranged from 22 to 30 years. All participants had perfect (natural or corrected) vision. Each session took about 60 minutes.

Design. Independent variables were the window size and the target distance. Since we varied the window size, we opted for a constant target width. We utilized two different window sizes, a small window with $5 \times 8\,cm$ (resembling the standard display form factor of today's smartphones like the Apple iPhone) and a larger window with $8 \times 8\,cm$. The dependent variable was the time it took a user to move the window from the center of the strip to the center of a target. The participants were asked to shortly confirm that they had reached the target. The participants used the apparatus horizontally on a table while being seated. They were seated at the center of the rail and the strip respectively. We did not mount the apparatus vertically to a wall to assess the performance while standing, since this only further limits the human motor system in the horizontal axis. We video-recorded the tasks and measured the navigation times manually by analyzing the videoframes.

We chose a within-subjects design, but split the participants into two groups with 16 participants each. Each group was assigned a different set of target distances. This allowed us to get a broader variety of target distances. The order of the tasks, target distances and data types was completely counter-balanced. The participants were introduced to the concepts and were allowed to familiarize themselves with the apparatus upfront.

Tasks and Target Types. We particularly wanted to assess the difference between the navigation in a known and an unknown information space. We therefore chose two different target types, *numbers* and *symbols* (see Fig. 3.16). The numbers were ordered naturally with 0 being at the center of the strip. They resemble a known information space, since users build a mental model of the number ray and map it to the strip easily. The symbols served as an unknown information space. The targets were distributed both equidistantly and non-equidistantly as shown in Figure 3.16.

The participants had to fulfill 16 tasks per window size (5 equidistant numbers, 3 non-equidistant numbers, 5 symbols and 3 non-equidistant symbols), resulting in a total of 32 tasks per participant and a total of 1024 data points. In case of the symbols, we showed the participants the symbol they had to look for before each task. The symbol remained visible throughout the whole task.

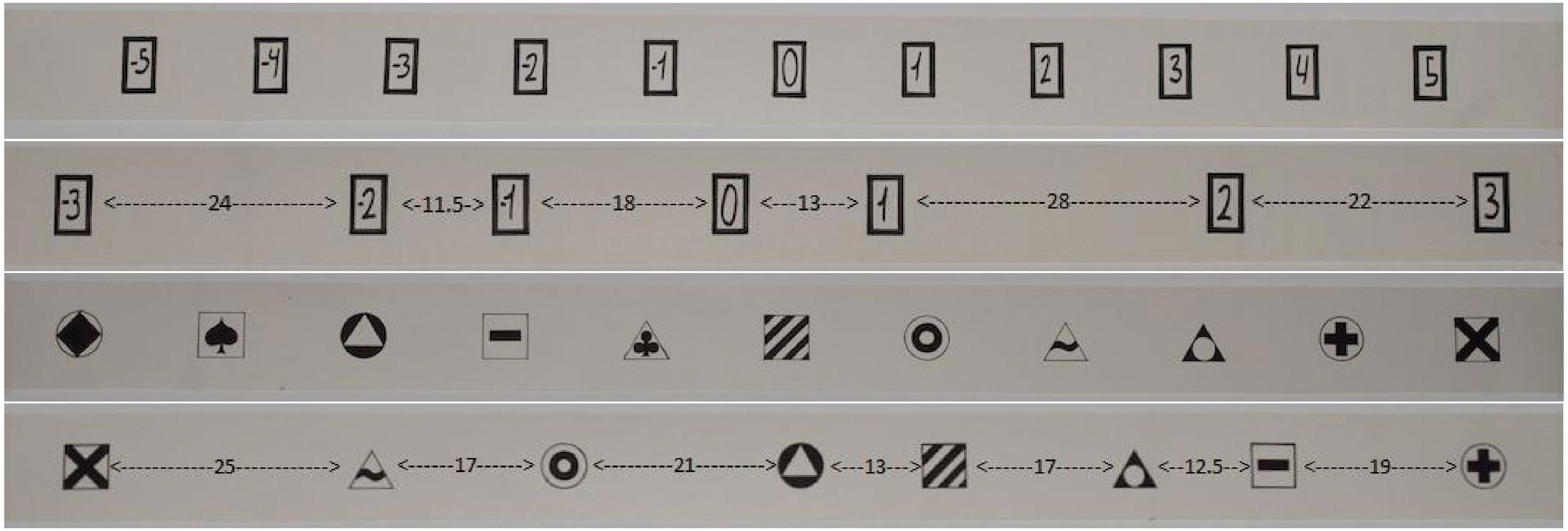


Figure 3.16: From top to bottom: (1) equidistant numbers, each 4 cm × 5 cm large and spaced 8.5 cm apart from each other, (2) non-equidistant numbers, (3) equidistant symbols, with a baseline of 4 cm and spaced 8.5 cm apart from each other, (4) non-equidistant symbols. The distances in the non-equidistant case were *not* printed onto the paper strip.

3.3.5 Results

3.3.5.1 Movement Time

The average movement times T per task are shown in Table 3.2. The movement time increased monotonically with the target distance for all data sets. ANOVA tests revealed that this effect is statistically significant ($p < .001$). Moreover, Bonferroni post-hoc tests confirmed that this holds for all distances ($p < .001$). Although the participants took longer in average to find a target using the small window and consequently were faster using the larger window, the speed difference was not significant for any of the data sets.

When the participants initially moved in the wrong direction, the movement time was significantly higher than when they directly moved towards the correct direction. ANOVA tests and Bonferroni post-hoc tests confirmed the significant effects for both equidistant symbols (small window: $F_{1,18} = 21.01, p < .001$; large window: $F_{1,18} = 15.39, p < .001$) and non-equidistant symbols (small window: $F_{1,10} = 30.82, p < .001$; large window: $F_{1,10} = 26.66, p < .001$). The statistically significant speed-up ($p < .001$) for equidistant symbols was 2.9 and 2.76 for non-equidistant symbols in average.

Data	Small		Large	
	T_{avg} [s]	SD_{avg}	T_{avg} [s]	SD_{avg}
Num-Equi	1.11	0.06	0.87	0.07
Num-NEqui	1.39	0.10	1.20	0.07
Correct Direction				
Sym-Equi	1.75	0.09	1.43	0.08
Sym-NEqui	1.82	0.08	1.40	0.06
Wrong Direction				
Sym-Equi	4.84	0.33	4.35	0.32
Sym-NEqui	4.53	0.36	3.86	0.38

Table 3.2: Average movement times per data set

3.3.5.2 Model Fitting

We fitted the movement times to the linear model from equation (3.2), Cao's model as in equation (3.1) and the trigonometric model from equation (3.4). Table 3.3 shows the parameter estimates, the respective standard errors for the estimates and the correlation coefficient R^2 for the numbers data sets; Table 3.4 for the symbols data sets respectively. Cao's formulae yielded the best fit for all number data sets, whereas our proposed model based on the tangent yielded the best fit for all symbol data sets. The tangent model fit particularly

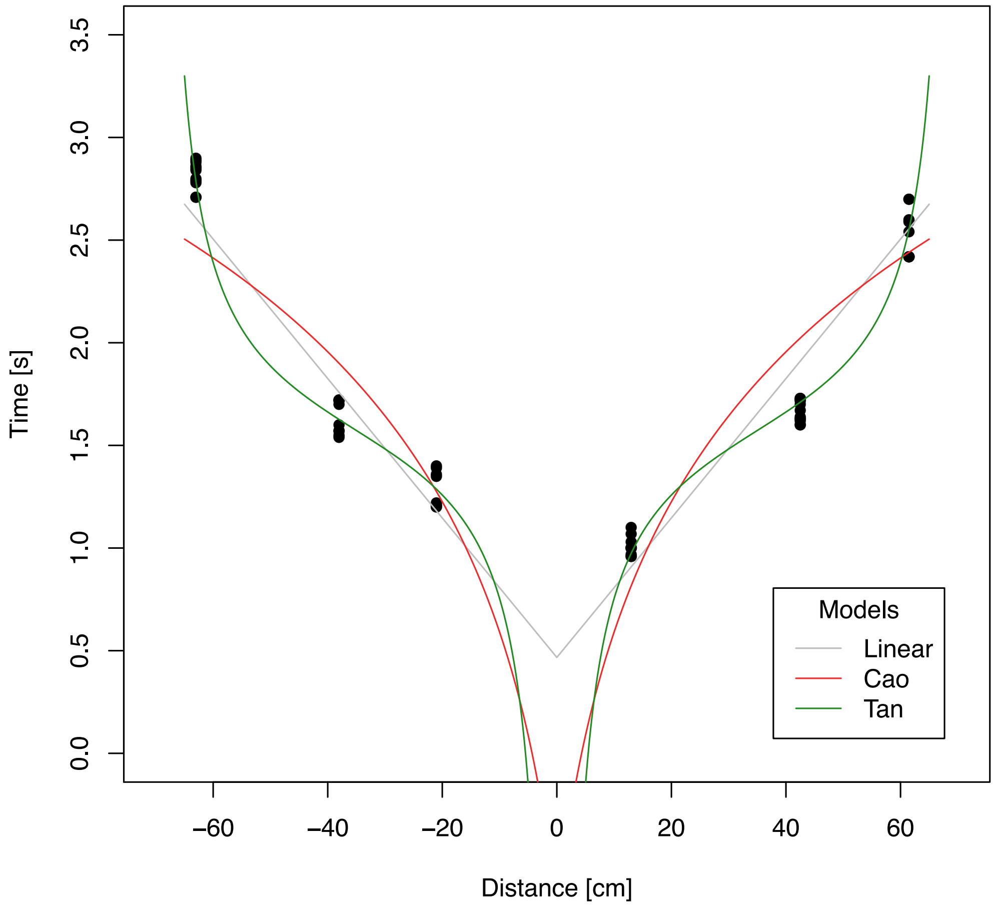


Figure 3.17: Model fits for non-equidistant symbols, small window. Participants initially chose the correct direction.

well for the case where the participants initially went in the wrong direction. The results are highly significant ($p < .001$). Figure 3.17 shows an example fit for all three formulae.

	Win.	F	a	E_a	b	E_b	R^2
Equi	S	Lin.	0.37	0.03	0.02	0.00	0.91
		Cao	-0.39	0.04	0.52	0.01	**0.94**
		Tan	1.12	0.01	0.31	0.01	0.92
	L	Lin.	0.42	0.03	0.01	0.00	0.86
		Cao	-0.62	0.29	0.16	0.08	**0.87**
		Tan	0.88	0.01	0.19	0.01	0.85
NEqui	S	Lin.	0.07	0.03	0.03	0.00	**0.98**
		Cao	-1.31	0.08	0.93	0.03	**0.98**
		Tan	1.29	0.03	0.41	0.02	0.90
	L	Lin.	0.18	0.02	0.03	0.00	**0.98**
		Cao	-0.65	0.05	0.79	0.02	**0.98**
		Tan	1.09	0.01	0.33	0.01	0.97

Table 3.3: Model fit data for the numbers data set.

			Correct Direction					Wrong Direction				
Data	Win.	F	a	Err_a	b	Err_b	R^2	a	Err_a	b	Err_b	R^2
Equi	S	Lin.	0.63	0.03	0.03	0.00	0.96	0.91	0.22	0.10	0.01	0.93
		Cao	-0.36	0.09	0.72	0.03	0.92	-3.11	0.54	2.69	0.17	0.89
		Tan	1.62	0.01	0.36	0.01	**0.98**	4.37	0.06	1.28	0.04	**0.98**
	L	Lin.	0.54	0.02	0.02	0.0	0.95	1.67	0.17	0.07	0.00	0.92
		Cao	-0.08	0.06	0.63	0.02	0.93	-0.35	0.38	2.02	0.16	0.87
		Tan	1.33	0.01	0.28	0.01	**0.97**	4.08	0.05	0.89	0.03	**0.97**
NEqui	S	Lin.	0.47	.06	0.03	0.00	0.96	2.73	0.18	0.05	0.00	0.88
		Cao	-0.77	0.15	0.86	0.05	0.91	1.02	0.39	1.16	0.13	0.83
		Tan	1.57	0.01	0.39	0.01	**0.99**	4.19	0.06	0.53	0.04	**0.92**
	L	Lin.	0.44	0.04	0.02	0.00	0.96	1.91	0.16	0.05	0.00	0.89
		Cao	-0.30	0.08	0.69	0.03	0.95	0.57	0.31	1.35	0.12	0.86
		Tan	1.22	0.01	0.28	0.01	**0.98**	3.48	0.06	0.57	0.03	**0.93**

Table 3.4: Model fit data for the symbols data set.

3.3.6 Discussion

The main goal of our experiment was to investigate how search times in unfamiliar information spaces can be modeled for embodied peephole interaction. Table 3.5 summarizes the model with its parameters.

We found that the movement times for embodied dynamic peephole navigation significantly depend upon the target distance (H3). While this might seem trivial, it has not been investigated with a physical apparatus before. We found that Cao's model fits best for a familiar information space such as a ray of numbers, as partly used in our experiment. Users can easily build a mental model thereof and the impact of the physiological constraints of the human body is only minor.

We also found a difference in the search times for the equidistantly and non-equidistantly distributed number rays. Although this difference is statistically not significant, it suggests that the user's familiarity with the information space has an impact on the navigation (which we claimed in H1). We were able to confirm this hypothesis with the results for a completely unfamiliar information space for which we used symbols. In this case, the actual movement of the window played an important role and thus the trigonometric model based on the tangent fitted significantly better (H2). The actual distribution of the symbols, whether aligned equidistantly or not, had no influence on the movement times. However, the average search times for the non-equidistant distribution of the symbols were lower than those for the equidistant distribution. We assume that this attributes to the information density in the information space. The strip with the non-equidistant symbols contained only 8 symbols, whereas the equidistant strip contained 11 symbols.

The size of the utilized window had no significant effect on the navigation times (H4). This leads to very interesting hypothesis with practical relevance: Since we chose window dimensions similar to the displays used in today's handheld devices, this might imply that designers can abstract from the actual display size to a certain extent when designing information spaces which are to be navigated with spatially aware handheld devices. Density and distribution of information elements in the information space can then be determined without having a concrete device in mind during the design phase. However, this remains to be investigated in future experiments.

Trigonometric Peephole Pointing Model	
$$T = a + \frac{b}{S} \tan\left(\frac{D}{L}\pi + \frac{\pi}{2}\right)$$	
$D \in \mathbb{R}$	Distance to target
$L \in \mathbb{R}$	Width of information space
$S \in \mathbb{R}$	Width of display
$a, b \in \mathbb{R}$	Empirically determined constants

Table 3.5: Summary of the peephole pointing model. The tangent is shifted by half a period and scaled by $\frac{D}{L}$ to match the interval of $(-\frac{L}{2}, \frac{L}{2})$.

The initial search direction had a significant impact on the actual navigation time (H5). When the participants initially moved in the wrong direction, the search task was prolonged by a factor of at least 2.5. In this case, the other models fitted even worse and the trigonometric model still fitted best. Although users typically have to decide where to move first in practice, previous research neglected this fact.

The participants used similar navigation strategies as observed by Cao et al. [2008]. Some participants employed a backtracking technique as well: they first moved the window quickly and hoped to get a glance at the target to backtrack it. However, this did not affect the search times significantly.

3.4 Conclusion

In this chapter, we addressed the mobile interaction with large information spaces from a *space-centric* interaction perspective. We extended the virtual information space to the physical space. For this purpose, we utilized mobile devices as spatially aware see-through displays, providing an embodied dynamic peephole onto the large virtual information space. This chapter therefore advances the field of mobile interaction with large multimedia information spaces with respect to the following contributions:

- **Empirically Grounded Theory for Embodied Peephole Interaction:** The analysis of prior work in this field showed that most of the approaches were evaluated in lab settings and not in a real world context, where they actually should be applied. Thus, the field lacked a fundamental understanding of how users would actually interact through embodied peepholes in mobile, real world settings. To address this shortcoming, we conducted a qualitative, exploratory field study. This allowed us to derive an empirically grounded theory, characterizing embodied peephole interaction in four inter-related categories. We found that the actual navigation is constrained by physical, as well as situational constrains and influenced by both visualization and user preferences.

- **Design Implications:** The inter-relationships of these categories allowed us to derive design implications for future embodied peephole interfaces. The results also revealed that the layout of the information space in physical space is highly important. Poorly positioned content might not be reachable. The layout can be optimized by modeling a user's performance through considering navigation times to targets depending e.g. on their distance.

- **Novel Movement Model for Embodied Peephole Navigation:** For this purpose, we contributed a novel model for the embodied navigation of a-priori unknown information spaces with spatially-aware displays. It models the navigation time to a target in one-dimensional information spaces, depending on the size of the display and the distance to the center of the target. The model is inspired by physiological aspects of the human body and thus particularly addresses the affordances of embodied interaction with such displays.

- **Empirical Model Validation:** The results of a controlled experiment with 32 participants using a physical apparatus validate our model for the navigation in a-priori unknown information spaces. We found that a user's familiarity with the information space and her initial search direction has a significant impact on the navigation time to hidden targets. When a user is familiar with the information space, the search times are logarithmically dependent as in Cao's case, whereas in the case of an unfamiliar information space, the search time significantly increases towards the physical end of the information space due to human physiology and is thus better modeled by a trigonometric formulae.

Our results provide hints that the density of the information space should be considered in future experiments. Our results furthermore lead to the hypothesis that the minimal variance in the display size of current handheld devices has no effect on navigation times. This should be investigated in future research, particularly in light of the growing importance of tablet devices such as Apple's iPad with varying form factors (i.e. display size, weight or look and feel). Furthermore, it is important to investigate how visual cueing–as discussed in the last section–can be used effectively to scaffold the navigation in unfamiliar information spaces. Last, future research should also focus on possible extensions toward the initially mentioned "full-fledged" model for pointing in 3D space.

CHAPTER 4

Integrating Device- and Space-Centric Interaction

The last two chapters approached the mobile interaction with large multimedia information spaces from two different perspectives. In Chapter 2, we showed how to design more usable and enjoyable interfaces for mobile devices despite their limiting device characteristics. Here, the user interface design focused particularly on the small screen and limited both information and interaction space to it, respectively. Chapter 3 then focused on pushing the boundaries of the virtual information space toward the physical space, therefore adopting a space-centric interaction perspective. For this purpose, we utilized mobile devices as spatially aware see-through displays, providing an embodied dynamic peephole onto the large virtual information space and thence creating a see-through augmented reality.

The present chapter integrates these two perspectives and at the same time transgresses the limitations of hand-held displays. We will investigate how virtual and physical space can be tightly integrated, allowing for both device- and space-centric interaction with real world objects–not just one particular mobile device. For this purpose, we propose to use pico projectors as so-called "light beams": everyday objects sojourning in a beam are turned into dedicated projection surfaces or tangible interaction devices. This way, our daily surroundings get populated with interactive objects, each one temporarily chartered with a dedicated sub-issue of pervasive interaction. In this chapter, we will explore the conceptual framework around *LightBeam*. Moreover, we will investigate how lightbeams can be used in real world settings through an exploratory field study. We will illustrate how these findings informed the design of novel interaction techniques. These techniques are specifically designed to (1) turn the draw-back of a small projection area into a benefit, (2) provide a trade-off between digitally augmented and traditional uses of everyday objects, and (3) work with almost any object within reach, which is important for nomadic settings. The techniques have in turn been implemented and evaluated in an early user feedback session, eventually.

The remainder of this chapter is structured as follows. In Section 4.1, we discuss challenges with pico projector interaction, particularly for interaction with real world objects. Moreover, we outline open research questions in this field. Section 4.2 presents the conceptual framework of LightBeam and relates it to prior research on pico projectors. Our analysis shows that the tangible character of real world objects has not yet been systematically explored for pico projector interaction. Moreover, it is unclear how the mobility of physical objects can be leveraged for tangible interaction. As a first step toward addressing these problems, we investigated these questions in an exploratory field study with interaction design researchers. Section 4.3 presents our exploratory field study and discusses our findings. The results provide detailed insights into the design space of projector-based tangible user interfaces with mobile real world objects. Based upon the qualitative results, we conceived and implemented various novel interaction techniques for 3D object interaction with pico projectors, which are illustrated in Section 4.4. These have in turn been evaluated in an early user feedback session. Section 4.6 reports on the results. The chapter concludes with a discussion and an outlook upon future work in Section 4.7. In summary, the main contributions of this chapter are

- conceptual framework for pico projector interaction,
- exploration of open research questions and empirically grounded theory based upon a qualitative, exploratory field study,
- novel interaction techniques for three-dimensional, spatial object interaction with pico projectors and
- evaluation of these techniques through early user feedback.

4.1 Overview

The capabilities of pico projectors have significantly increased lately. In addition to their portability, they allow us to dynamically project digital artifacts into the real world. Since pico projectors have been around for some years now, there is a growing body of research on how they could be integrated into everyday workflows and practices. Two major categories of corresponding interaction techniques have evolved [Rukzio et al., 2011, Cowan et al., 2012]: (1) using the projector itself for input (either via direct input such as buttons on the projector or by moving the projector like a flashlight); (2) interacting on the projection surface via direct touch or pen-based input. The projection surface is usually supposed to be fixed, large, and flat.

The present chapter investigates pico projectors for interaction with real world objects–which is fundamentally different: when we engage with real world objects such as physical paper or a coffee mug, we move the objects in three dimensions and engage with them spatially: we pass a piece of paper to a colleague, we lift the coffee mug to take a sip, etc. This is particularly interesting considering recent technological developments which integrate mobile phones and projectors. Smartphones with built-in projectors will influence or even determine how projectors are used in our everyday activities. Instead of being held in hand all the time, mobile phones are often placed onto tables, for instance during meetings. Thus physical objects on the table move into the projector's reach (cf. Figure 4.1). This enables a novel kind of interactive tabletop: not only the table surface, but the objects on the table become interactive displays. Intuitive handling of such objects has the potential to foster rich, non-obtrusive and tangible UIs.

This chapter presents a novel interaction concept for pico projectors and real world objects, which we call *LightBeam*. The LightBeam metaphor leverages real world objects as projection surfaces when brought into the projection beam; spatial manipulation of the objects is interpreted as user input and influences the projected content. We tend to think of this kind of interaction as a third stage of *pervasive display-centered interaction*, the first stage being ubiquitous availability of interactive displays (smartphones and touch screens everywhere), the second stage being ordinary flat surfaces combined with pico

Figure 4.1: Pico projector is placed on a table and uses a nearby espresso cup to show email notifications (concept).

projectors and direct manipulation input (touch, pen, etc.). In the third stage considered here, arbitrary objects become display surfaces; at the same time, the content displayed and the interaction concepts become object specific. Additional objects brought into the projection ray correspond to additional projection surfaces, adding another degree of freedom, e.g. for tangible interaction. These observations lead us to the following research questions:

> How can three-dimensional, physical objects be used for interaction in combination with pico projections in nomadic settings?
>
> What type of digital information should be displayed on which kind of objects?
>
> How to cope with the very limited field of projection?

Before addressing these questions through an exploratory field study with interaction design researchers in section 4.3, we first present the conceptual framework of LightBeam and relate it to prior research on pico projectors in the next section.

4.2 Related Work and Conceptual Framework

There is already a notable body of knowledge on pico projector interaction. Figure 4.2 shows the conceptual categories for this kind of interaction. We will discuss both background and conceptual framework of LightBeam in the context of these three categories.

4.2.1 Fixed Projector & Fixed Surface

The small form factor of pico projectors can be leveraged for integrating them virtually anywhere. In *Bon re* [Kane et al., 2009], two camera-projector-units are attached to a laptop and therefore extend the display area to the left and right hand sides of the laptop. The projection is used as an interactive surface, allowing users to employ multi-touch gestures on the projected area. Moreover, the system recognizes everyday objects such as a coffee cup through vision-based methods and can project additional information, however only onto the flat table surface, not onto 3D objects. *FACT* [Liao, Tang, Liu, Chiu, and Chen, 2010] tracks ordinary paper documents with their natural features and enables word-level augmented reality interaction with the documents. Both projector and paper document need to be placed at a fixed position to enable fine-grained document interaction. Other examples are indirect input techniques using gestures [Cauchard et al., 2012] or shadows [Cowan and Li, 2011].

The aforementioned research focuses on techniques, where both projector and projection surface are required to be fixed in space (cf. Figure 4.2a).

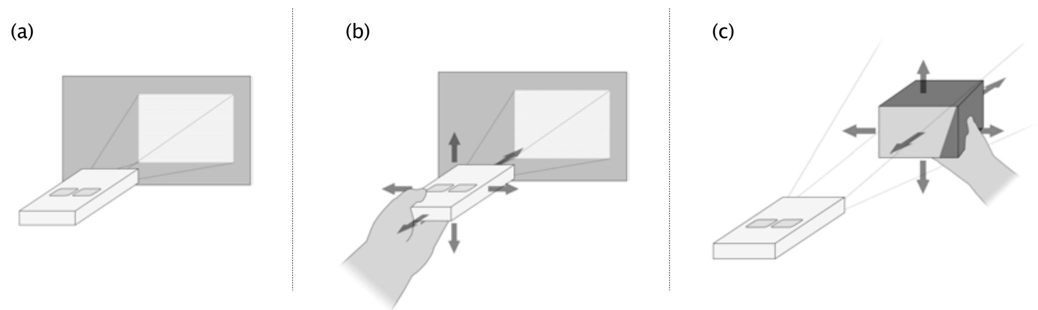


Figure 4.2: Conceptual levels for pico projector interaction: (a) fixed projector, fixed surface; (b) mobile projector, fixed surface; (c) fixed projector, mobile surface (LightBeam–the beam is used for *output on* as well as *input with* physical objects sojourning therein).

4.2.2 Mobile Projector & Fixed Surface

A larger body of research is motivated by the mobility of pico projectors [Wilson, Robinson, Craggs, Brimble, and Jones, 2010]: they can be easily carried around, held in hand and project onto fixed surfaces such as walls (cf. Figure 4.2b). Prominent work has been carried out by Cao et al. [Cao and Balakrishnan, 2006, Cao et al., 2007]. They developed various handheld interaction techniques, as well as pen-based techniques for direct input on the projection surface. In both cases, they chose large flat and fixed surfaces, such as walls, as their projection targets. Most of the techniques rely on the so-called flashlight metaphor. Similar to peephole interaction as discussed in chapter 2, the projector only projects a subpart of the virtual information space. By moving the projector, further parts of the information space are being revealed. The flashlight metaphor is also used in other projects such as MapTorchlight [Schöning et al., 2009], Marauder's Light [Löchtefeld et al., 2009], iLamps [Raskar et al., 2006], RFIG Lamps [Raskar et al., 2004], MouseLight [Song et al., 2010] and PenLight [Song et al., 2009] to augment static surfaces with digital information. The latter two also allow for direct pen input on the projection surface. Most recently, Molyneaux et al. [2012] have presented two camera-projector systems, which support direct touch and mid-air gestures on arbitrary surfaces. However, once registered, these surfaces must remain at a fix location, which impedes tangible interaction. MotionBeam, a concept by Willis, Poupyrev, and Shiratori [2011b], equally uses a fixed surface as projection target. It allows users to steer a projected virtual character through virtual worlds. The character is bound to the projection; the projector is handheld and reveals only a part of the game world. By moving the projector, users can dynamically reveal further parts of the virtual world. Willis et al. have also contributed SideBySide [Willis, Poupyrev, Hudson, and Mahler, 2011a], which focuses on ad-hoc multi-user interaction with handheld projectors on fixed surfaces.

A few projects also investigated wearable projection, where the pico projector is attached to clothes or worn like an accessory. A prominent example here is Sixth Sense [Karitsuka and Sato, 2003]. A camera-projector unit is worn as a necklace. Physical surfaces such as walls, but also parts of the body can then be used as a projection surface. Users are able to interact with the projection using in-the-air gestures in front of the camera. Skinput [Harrison, Tan, and

Morris, 2010] also leverages body parts as projection surfaces but allows for touch input directly on the body. This effort has been further refined in OmniTouch, where Harrison, Benko, and Wilson [2011] enabled touch input on arbitrary surfaces using a depth-camera and a pico projector. Although these three projects support projection onto essentially mobile objects such as a human arm, these objects are only used as hosts for the projection, not for tangible interaction. Hence, from a conceptual viewpoint, they can also be regarded as fixed projection surfaces. A slightly different approach is pursued in Cobra [Ye and Khalid, 2010] by Ye and Khalid. They use a flexible cardboard interface in combination with a shoulder-mounted projector. The cardboard can be bent as a tangible input for mobile gaming but needs to be held at a fixed position.

4.2.3 Fixed Projector & Mobile Surface

In summary, previous work on pico projector interaction emphasized fixed and flat projection surfaces in physical space. The interaction was either projector-centered or relied on direct input on the projection surface.

It is worthwhile to note that there is a larger body of knowledge on projection-based interaction with larger projectors. Prior work in this field dates back to the early 1980s, when Michael Naimark investigated immersive projection environments in art installations [Naimark, 2005]. More recently, physical objects such as paper have been used as projection surfaces in *PaperWindows* Holman et al. [2005]. This very idea has been developed further in *LightSpace* [Wilson and Benko, 2010], where basically any *xed* surface in a small room installation is being recognized. Within this scope, Wilson et al. have investigated interaction on, above and between surfaces–but not with the surfaces themselves as e.g. tangible interaction devices. Most related to our work is Molyneaux's work on smart objects [Molyneaux et al., 2008, Molyneaux and Gellersen, 2009]. They have investigated how physical objects can be turned into interactive projected displays. The main focus of the work was on orchestrating a technical infrastructure, allowing for a reliable and robust object detection through model-based approaches. In addition to relying on larger projectors, they have not investigated the tangible character of physical objects, but used the projections to display additional object-specific information directly on the objects.

However, compared to larger projectors, the affordances of pico projectors are fundamentally different: they are mobile and have a very small and strictly limited projection ray. Thus we tend to think of pico projectors more like personal devices, which are carried around in a nomadic way during the day and used in a plethora of situations and places, such as workplaces or cafés. To the best of our knowledge, the tangible character of real world objects has not yet been systematically explored for pico projector interaction. Moreover, it is unclear how the mobility of physical objects can be leveraged for tangible interaction and what kind of projected information matches the affordances of physical objects (cf. Figure 4.2c). LightBeam aims at filling this void.

4.2.4 Our LightBeam Concept

In LightBeam, the pico projector is placed in the vicinity of the user and not constantly held in hand. It can be attached to physical objects (e.g. walls, desks or cupboards) and its tilting angle can be adjusted. This way, projection onto the physical space can be supported from flexible perspectives. Figure 4.2c illustrates the LightBeam concept. The projection is regarded as a constant ray of light into the physical space. The projection is "always-on" during an entire session of usage. The projector itself is augmented with a camera-unit and can track objects within its ray in three-dimensional space. Thus, the beam is bi-directional in nature, providing both output and input functionality:

- **Output-Beam:** Object surfaces can be (1) capitalized as **display** surfaces or (2) **augmented** with contextually meaningful digital information or functionality relevant to the object used as projection target.

- **Input-Beam:** Deliberately moving an object into the ray and manipulating it there can also serve as **input** (e.g. on the projection surface or through tangible interaction with the object itself).

For instance, a physical document held into the ray can be automatically recognized and contextually relevant information can be displayed on the physical document (*augmentation*). Moreover, physical interaction with the objects such as movement, rotation or other embodied gestures can be used as tangible control. For instance by gradually bringing a document into the ray (*input*), the level of detail of projected contents is continuously increased (*display*).

Central to LightBeam is the concept of moving objects but not the pico projector (except for changing the perspective). Thus, LightBeam provides a theoretically motivated conceptual framework, focusing on object-centered, spatial interaction and a three-dimensional projection space.

Figure 4.2 separates the composition of projector and object mobility conceptually. In practice, the boundaries are not rigid and the individual approaches can be combined, leading also to mobile projector interaction with mobile objects as a combination of Figure 4.2b) and 4.2c).

4.3 Exploratory Field Study

We have setup the conceptual framework for LightBeam and related it to prior research. As a second step, we conducted an exploratory field study to investigate the aforementioned research questions and to gain a deeper understanding of how pico projectors can be used together with physical objects in the context of LightBeam. Besides exploring the design space, the qualitative results should also inform novel interaction designs. We particularly wanted to explore the following dimensions:

- **Projector placement:** How is the projector positioned in physical space? For instance, is it hand-held or is the projector deliberately placed in the environment?
- **Output:** What kinds of objects are used for mobile projection? What kind of information should be displayed, depending on the target objects? Does mobile projection influence the dedicated meaning of objects?
- **Input:** How are real world objects manipulated in 3D space for interaction with mobile projections?

In the following, we outline our study design, the employed methodology and discuss the findings in detail.

4.3.1 Study Design

4.3.1.1 Setting

We conducted the study in two different places with each subject: the subject's workplace and a café. We selected these two places mainly for three aspects: spatial framing, social framing and the manifold nature of objects contained within these places. In particular, these places allowed us to study personal places, which are thoughtfully arranged by the participant and contain personal objects, and public places, where available objects typically do not have a personal meaning to the participants. Figure 4.3 shows examples of both places. For the café setting, we ensured that the types of objects present on the coffee table were consistent for all sessions. This was not desired for the office setting, since it was the subject's personal desk. The participants were seated in both settings. Each session lasted about 1.5 hours in average. The order of the places was counter-balanced.

4.3.1.2 Participants and Tasks

We recruited 8 interaction design researchers (7 male, 1 female) between 25 and 33 years of age (mean 28). Their working experience as interaction design researchers ranged from 1 to 6 years (mean 4 years). All participants were owners of fairly modern smartphones with touch interfaces such as the iPhone. Moreover, all participants were familiar with novel input techniques such as physical interaction (e.g. tilt-based input using an accelerometer) and gesture-based interaction (e.g. touch input on an interactive surface like a tabletop).

Our main objective was to observe the participants while using the projector for certain interactions in the field. The interactions themselves were embedded in semi-structured interviews. The participant was given an Aaxa L1 laser pico projector and plenty of time for getting familiar with the pico projector.

The participants were told that the projector could be used for the same tasks as they carry out with their mobile phone. The projector was able to display a number of multimedia resources such as photos, videos and digital documents that we had selected and stored on the device before. The content was used during the sessions to simulate typical scenarios for pico projector usage such as photo sharing, video consumption or co-located collaboration with

Figure 4.3: Example photographs from the two settings in the exploratory field study; personal desk (left) and café (right).

digital documents. The participants were either asked how they would project and interact with certain content or deliberately confronted with a projection. Figure 4.4 shows the latter case, where the interviewer projected a movie onto a cup on the participant's personal desk. The interviewer first observed how the participant would react to this and then continued the interview process. The semi-structured interviews were highly interactive and had the character of brainstorming sessions.

We used an Aaxa L1 laser pico projector, as a low-fidelity prototype. This was due to two reasons: (1) we did not want to influence the participants by any design and (2) we wanted to explore the aforementioned fundamental dimensions such as projector placement. A high-fidelity prototype would have imposed too many constraints on the interaction space.

4.3.1.3 Data Gathering and Analysis

We chose a qualitative data gathering and analysis methodology, which we performed iteratively per session. As data gathering methodologies, we used semi-structured interviews, observation and photo documentation. After each session, the interviews and observations were transcribed. Salient quotes were selected and analyzed using an open, axial and selective coding approach [Strauss and Corbin, 2008]. The emerging categories served as direct input for the follow-up session with the next participant. The scope of the session was adapted according to the theoretical saturation of the categories.

In the following three subsections, we present the findings from our study. The coding process yielded various categories, depending on where the projector was placed, which objects were selected as projection targets and how objects actually foster input capabilities.

4.3.2 Results I: Handheld versus Fixed Projector

Our observations revealed that the projector was used in a two-step process by all participants in both settings (office and café): initially, the participants used the projector as a handheld device to find a suitable projection area for the beam, which is not physically constrained by objects that cannot be moved. Then, they placed it onto the table and the projector was no longer used in hand throughout the entire session. The only exceptions were rare cases when the projector was moved to another location in its vicinity to slightly readjust the projection space.

Figure 4.4: Projection of a YouTube clip on a coffee mug.

Placing the projector instead of using it in hand was mostly due to ergonomic reasons. Once the projector was placed on the table, not the projector, but movable objects were repositioned to serve as projection targets. P8 noted: "*When would I actually make the e ort of holding the projector? I am constantly looking for objects, which are perfect hosts for the projection, which I can then bring into the beam. I do not want to hold the projector. It constrains me.*"

4.3.3 Results II: How to Leverage Objects for Output?

In the interviews, the participants noted that the affordances of objects determine whether and how an object can be used for output or input.

4.3.3.1 Relationship between Projected Content and Object

We observed a direct correspondence between the cognitive demand required by the projected content and both the size and shape of an object that was chosen as the projection target.

Cognitively demanding content such as presentation slides, where it is crucial to grasp the whole level of detail, was projected onto larger, less mobile and rigid surfaces. Examples comprise larger boxes, tables or the floor. Interestingly, such content was not projected onto walls, since in this case others would have been able to see it. The latter was considered either "*impolite and a disturbance to others*" (P5) or a privacy issue (mentioned by all participants).

Cognitively less demanding content, such as short YouTube clips or photos, was projected onto rather small and even non-planar objects (e.g. see Figure 4.4). Participants commented that these are perfectly suitable when only a lower level of detail is required. Moreover, such objects provide the benefits of being easily movable. As a direct consequence, they can be easily replaced by other objects when required. For instance, P8 used the back of his hand as a substitute projection surface, when he viewed a projection together with the interviewer and was required to move the original surface (a rigid paper box) away. He stated: "*I considered it impolite to just leave you without the projection. So I gured out that the back of my hand is better than nothing at least you can see the projection*".

The participants did not mind slightly distorted projections, when they did not want to devote their whole attention to the projection: "*I do not care that this projection [a YouTube clip] does not t onto this object [a small package, 5x3cm] I still can understand the gist of it*". Moreover, even curved surfaces were used for such a task, e.g. P7 commented in the situation of Figure 4: "*Even though it is distorted towards the edges of the cup, I do not mind, since it is not a high quality movie. Moreover, I only focus on the center of the projection and I can understand what is actually happening*".

4.3.3.2 Objects afford Physical Framing

The natural constraints provided through the boundaries of physical objects were also considered important. P7 noted: "*I want to put things into frames. Objects on my desk provide this frame, whereas my table itself is too large there is no framing*". This is different to just projecting a digital frame around the projection, since moving the frame would imply moving the projector. But here, objects are the frames. It was considered crucial that the projection is clearly mapped to the object. P8 elaborates on this by saying: "*Objects are like frames for me, they provide space and receive the projection*".

4.3.3.3 Embodiment of Digital Artifacts

We observed that all of the participants used the mobility of objects and the physical framing of the projections to control who is actually able to see the projected content. P2 stated: "*You can easily direct attention by moving it, [turns a menu with the projection on it to herself] and now I can read it.*" This leads to a rather object-centric perspective on interaction, as P3 outlines: "*It is not the device I care about, it is the object with the projection.*" Moreover, P4 argues that "*the data is on the object, it is contained within it. The digital artifact is embodied through the physical object.*"

4.3.4 Results III: How to Provide Input with Objects?

While larger surfaces provide extensive display area for detailed output, they are likely hard to move and therefore are rather fixed in physical space. Smaller physical objects however afford manipulation in three-dimensional space.

4.3.4.1 Moving Objects within the Beam

The participants argued that since the data is bound to a physical object, the object itself could be used as a tangible control. P7 described this as "*physical shortcuts to certain digital functionality*". He further mentioned that he makes "*an abstraction from the actual object towards its Geometry*". He therefore concludes: "*For instance, when I look at my co ee mug, I see an object which can be rotated by grabbing its handle; I would want to use this for quickly controlling something like a selection*". Another participant moved his hand forth and back within the projection ray and imagined to quickly skim through a list of pictures (cf. Figure 4.5). P6 noted that he "*would not want to perform a three-dimensional gesture mid-air due to the lack of haptic appeal, but using an object for that as a medium would be perfectly ne*".

Figure 4.5: A participant demonstrates how he would use his hand to quickly skim through a list of pictures and then turn his hand towards the interviewer to present a selected picture.

4.3.4.2 Dynamic Modification of Object Shapes

The flexibility that some physical objects exhibit, such as paper, was also used to dynamically modify the projection surface in two ways: (1) to increase and decrease the display size and (2) for (semantic) zooming, comparable to tangible magic lenses [Spindler, Tominski, Schumann, and Dachselt, 2010], but in a mobile situation. Participants used folding gestures with paper to increase or decrease the display size. Folding paper was mapped to decreasing and unfolding paper was mapped to increasing display size.

Participants reported that deformable objects are perfectly suitable for "*taking a peek into the beam*" (P5). P5 imagined that the projector was constantly projecting into space without a target object and was able to display notifications, like on his Android smartphone. "*By lifting a paper and moving it into the beam*", he explained, "*I can just take a look at my noti cations, you know, to look if something is there*".

4.3.4.3 Capturing Objects Visually

In the context of document interaction, the projector was also considered as a "scanner". P7 stated: "*If I project onto a document, the projector can also copy the physical document to the digital world. I can do this with various documents on the go and share them here.*" P2 also noted that the mobile projection can be used to add digital artifacts such as annotations to documents. She exemplified this by lifting an article, grabbing a pen and circling a paragraph.

4.3.4.4 Overloading Mappings of Physical Objects

Projecting onto an everyday object and mapping digital functionality to it is more than just a visual overlay in physical space. It also redefines the object's purpose. Moreover, a projection locks objects in physical space, as P7 elaborates: "*If I used this co ee mug as a tangible control for an interaction I heavily rely on, I would certainly have to forget its use as a mug. It would have to remain there, at that very place, to allow me to carry out this function at any time.*" The consensus across the participants was that overloading the mapping of physical objects is good, for short terms. Physical objects afford casual interaction, as P5 described: "*I would want to just put the object within the beam, carry out an interaction and remove the object from the beam*".

4.3.5 Summary

The results from our exploratory field study are summarized in Table 4.1. These show that LightBeam provides a fundamentally different interaction space for tangible interaction than larger immersive projection spaces. Being placed in a user's vicinity, it provides a dedicated interaction space through its highly limited projection ray. Our results show that moving objects therein is a central theme for interaction in real world settings. Objects provide a physical framing for projections and thereby embody them. Different physical characteristics of objects afford for projecting different digital contents. Furthermore, our results show that LightBeam, as a spatial ray, is not only used for output or tangible interaction, but also for capturing physical objects visually.

Projector Placement	× Projector is used in a two-step process: 1. Find suitable projection area with projector in hand. 2. Place projector in the vicinity (e.g. on the table). × Movable objects are repositioned to serve as projection targets.
Output	× Larger, less mobile objects afford cognitively demanding content. × Small and even non-planar objects afford less demanding content. × Objects afford physical framing of digital artifacts. × Embodiment of artifacts supports object-centric interaction.
Input	× Moving objects within the beam can be used for tangible control. × Generic object affordances can be leveraged for interaction. × Overloading mappings of physical objects is good–for short terms. × Dynamic modification of object shapes affords • flexible increase and decrease of display sizes, • adaptive level of detail ("to take a peek into the beam"). × Projector beam considered to capture objects visually.

Table 4.1: Summary of the results according to the initial research dimensions.

4.4 Interaction Techniques

Based upon our observations above, we have identified interaction primitives for LightBeam. These serve as the basis for interaction techniques discussed afterwards.

4.4.1 Interaction Primitives

We distinguish between the following interaction primitives (cf. Figure 4.6).

Move into the beam: Physical objects can be moved into the beam. In addition to moving an object entirely into the beam, the user can vary the degree to which the object re-sides within the beam. The portion of the object, which is located within the beam can be augmented with digital functionality. Several objects can reside simultaneously within the beam.

Remove from the beam: Removing an object from the beam removes any digital functionality from the physical object.

Move within the beam: Objects can be moved within the beam in three-dimensional space. This can be used to arrange projected contents in 3d space or as tangible control.

Beam captures an object: A visual copy of a physical object in the beam is captured and stored digitally.

Externalizing captured objects: Previously captured copies of objects can be visualized within the beam by projecting them onto physical objects.

In the following, we show how combining these primitive interactions creates novel interaction techniques that leverage the limited projection ray of LightBeam. We identified two promising application scenarios: on the one hand, when placing the pico projector on a table (similarly to how many people put their smartphones on a table during a conversation), it can turn everyday objects in its vicinity into peripheral awareness devices. On the other hand, LightBeam can aid in bridging the digital-physical divide when interacting with paper documents, a class of physical objects that is specific due to its high information content.

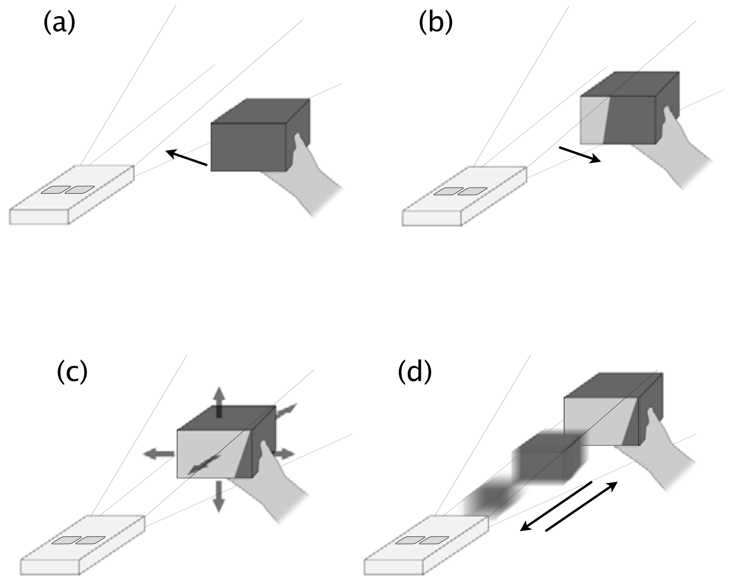


Figure 4.6: Interaction primitives for LightBeam: (a) Move into the beam, (b) Remove from the beam, (c) Move within the beam, (d1) Beam captures an object (direction toward projector) and (d2) Externalizing captured objects (direction toward object).

4.4.2 Gradual Sneak-Peek Into the Beam

Easily movable objects can be used to display information in-situ by moving them into the beam. Different objects afford different levels of details: while a larger box placed within the beam can show richer information (cf. Figure 4.7), smaller objects, e.g. a corner of a piece of paper, afford peeking at low-level information notifications.

We leverage the restricted field of projection for quick transitions between different levels of details. As an object is gradually moved into the beam, the projection area increases and more information can be presented. By partially removing the object from the beam, the level of detail of the information presented decreases. While this interaction is possible with any object, we believe that deformable objects lend themselves particularly to this interaction:

Figure 4.7.1 shows our exemplary interface: the projector is placed on a desk while the user is working with a physical document. The sketched projection ray in figure 4.7 indicates the highly limited projection area. The dotted line designates the effective projection (EP) area, which is the intersection be-

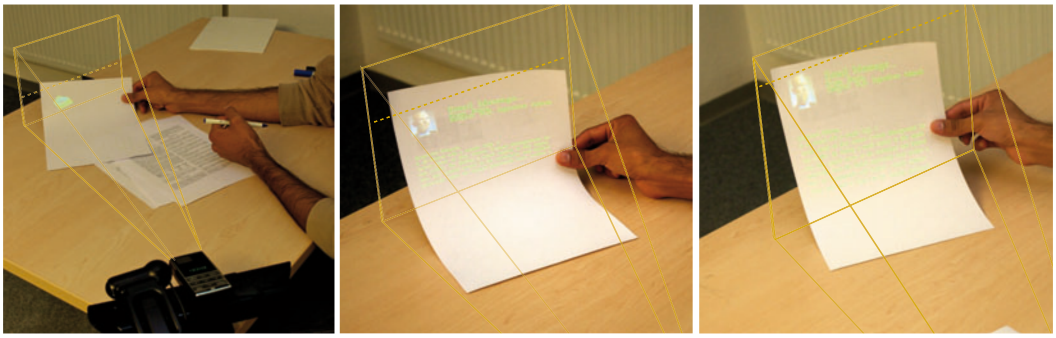


Figure 4.7: From left to right: the user utilizes the back of one of the papers he is currently working on to take a quick look into the projector beam. In the first image, a small envelope is displayed due to the limited projection space. By gradually lifting the paper, the level of detail is adjusted, more text is displayed and automatically wrapped within the boundaries.

tween the projection area and the object. By slightly lifting the document, the user can take a peek into the beam (small EP) and see if there are any new notifications. Gradually lifting the document further into the beam reveals more details (larger EP, cf. Fig. 4.7.2 and 4.7.3). Removing the paper from the beam reduces the EP and displays less information. As a slight variation of this technique, folding and unfolding a piece of paper within the projection beam affords a discrete transition between different levels of detail. As a matter of course, objects can also be permanently placed within the beam to immediately receive notifications (push-mode instead of pull-mode of information updates).

Projected contents can be bound to objects of particular shape (e.g. boxes as large displays as in Fig. 4.8). Alternatively, depending on the application or user preferences, contents can also be displayed on any object that is introduced into the beam. This ensures high usability in mobile contexts where specific objects might not be always at hand.

4.4.3 Using Any Object as Tangible Control

When moved within the beam, objects can act as tangible controls. Prior work [Cheng, Liang, Chen, Laing, and Kuo, 2010] mapped one particular object to a specific digital functionality. However, in nomadic settings, it cannot be taken for granted that specific objects are always available. Therefore, we advocate mapping a specific function not to one specific object, but to a class of objects that have a certain affordance. For instance, a function could be mapped to the physical rotation of a cylinder; hence any cylindrical object that affords rotation can be used to perform that function, e.g. a mug, a bottle, a vase, or a candy box.

Our implementation is shown in Figure 4.8. We use the rotation of objects, here a mug, to navigate through the displayed pictures. In particular, a physical object is only mapped to digital functionality while residing within the limited beam. Removing the object from the beam also removes the digital functionality and its original mapping is restored. Putting objects into the beam and removing them from the beam provides a lightweight way for switching between their uses as non-augmented vs. digitally augmented objects. For instance, when the coffee mug is not inside the beam, the user can take a sip from the mug without the system detecting this as tangible input.

Figure 4.8: A photostream from Flickr is projected onto a box and can be navigated by rotating the coffee mug.

4.4.4 Using the Beam as a Visual Scanner

In addition to projecting visual output onto objects or leveraging them as tangible controls, the beam can also be interpreted as a visual scanner, which captures objects. Moving an object into the beam selects it for capturing. Figure 4.9.1 and 4.9.2 show an example where a physical document is captured, automatically identified and its digital representation (here: a PDF) is stored virtually. With this technique, multiple pages (or documents) can be scanned subsequently. We model the process of capturing multiple objects as putting them onto a virtual stack of objects that resides within the beam: each scanned object is put onto the beam's internal stack and is stored digitally. The digital versions can in turn be externalized into the physical space by moving an object into the beam. Moving the object back and forth within the beam (see Fig. 4.9.3) allows for browsing the beam's stack. As an extension to the original "put-that-there" metaphor [Bolt, 1980], the beam's internal stack can also be transferred to nearby displays by a pointing the LightBeam toward a display to "beam-that-there".

Instead of scanning each object in its entirety, we also support more fine-grained selection. Figure 4.10.1 shows an example where a physical document is moved into the beam. In addition, a pen is also moved into the beam and can be used for selecting parts of the documents for capturing. Only selected parts are put onto the beam's stack. After the paragraph has been marked, its digital equivalent is copied, projected and can be moved in 3D space (cf. Fig. 4.10.2) using the pen.

In the reverse direction, the pen can be also used for putting a document snippet, which was previously captured by the beam, to a specific location on an object (the same object it was captured from or a different object). This is performed by a flick gesture with the pen towards the object (cf. fig. 4.10.3). There, the projection can be for instance used for spatial comparison (cf. Fig. 4.10.4). As described above for tangible interaction, the mapping of the pen is only temporarily overloaded. Moving the pen into the beam allows using it for copy and paste of document snippets. In turn, removing it from the beam restores its original function: it can be used for writing.

For the sake of focus and clarity, we here concentrate on tangible, ray-based interaction techniques. As a matter of fact, they can be easily combined with touch input, using the approach presented by Harrison et al. [2011].

Table 4.2 summarizes the interaction techniques and illustrates the interrelationship between the techniques, the utilized interaction primitives and the identified and partly refined concepts from the exploratory study.

	Primitives	Concepts	Techniques
Gradual Sneak-Peek Into the Beam			
Output	move into / out	× Physical framing of digital artifacts × Embodiment of artifacts	In-situ information display (i) Pull: move object into the beam, then withdraw it (ii) Push: object remains in the beam
Input	move within	× Dynamic modification of object shapes	(i) Flexible increase and decrease of display sizes (e.g. folding) (ii) Adaptive detail (e.g. flexible shape)
Using Any Object as Tangible Control			
Output	move into / out	× Overload mapping of physical object and augment it with functionality	Lightweight switching: objects within the beam are augmented. Augmentation is dissolved when moved out.
Input	move within	× Leverage generic object affordances for interaction × Move object within the beam for tangible control	Every object which has a certain affordance is mapped to a particular digital function, e.g. rotation is mapped to scrolling.
Using the Beam as a Visual Scanner			
Input & Output	capture	× Capture objects visually × Fine-grained capture	(i) Moving an object into the beam selects it for capturing. The object is then stored digitally on the beam's stack. (ii) Additional objects can be used for fine-grained selection (e.g. a pen).
Input & Output	externalize	× Browse captured objects × Externalize objects	(i) Moving objects within the beam browses visually captured objects. (ii) Captured objects can be externalized by gestures (e.g. flicking) or beaming (e.g. "beam-that-there")

Table 4.2: Interrelation of interaction primitives, concepts and interaction techniques.

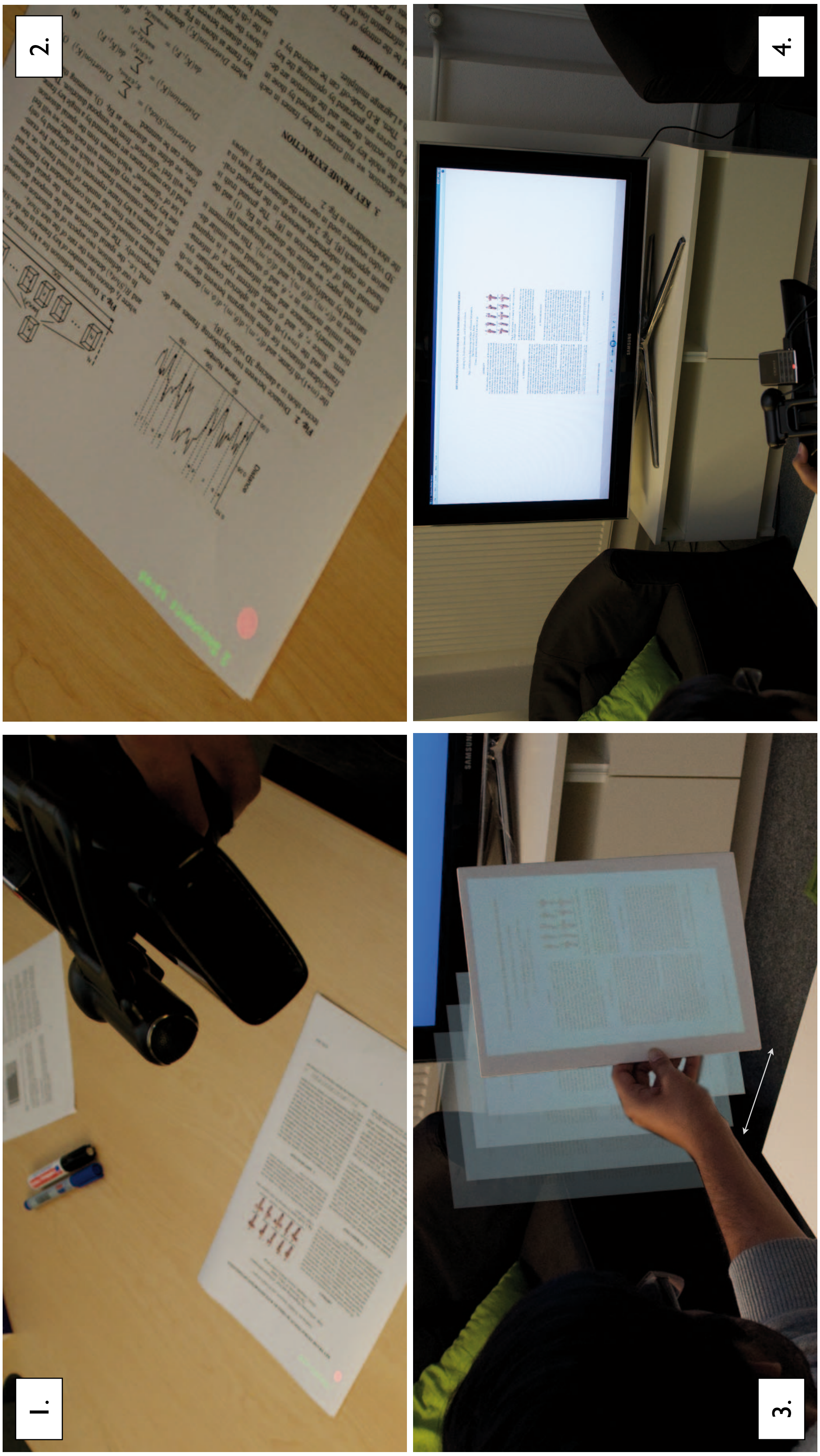


Figure 4.9: From left to right: (1) and (2) the LightBeam is used to recognize a physical document, storing its digital equivalent as a PDF. (3) shows a user skimming through a stack of captured documents by moving a piece of paper. Last, (4) shows the "Beam-that-there" technique. The user points the LightBeam to a TV in the vicinity and beams the documents on the stack to that very display.

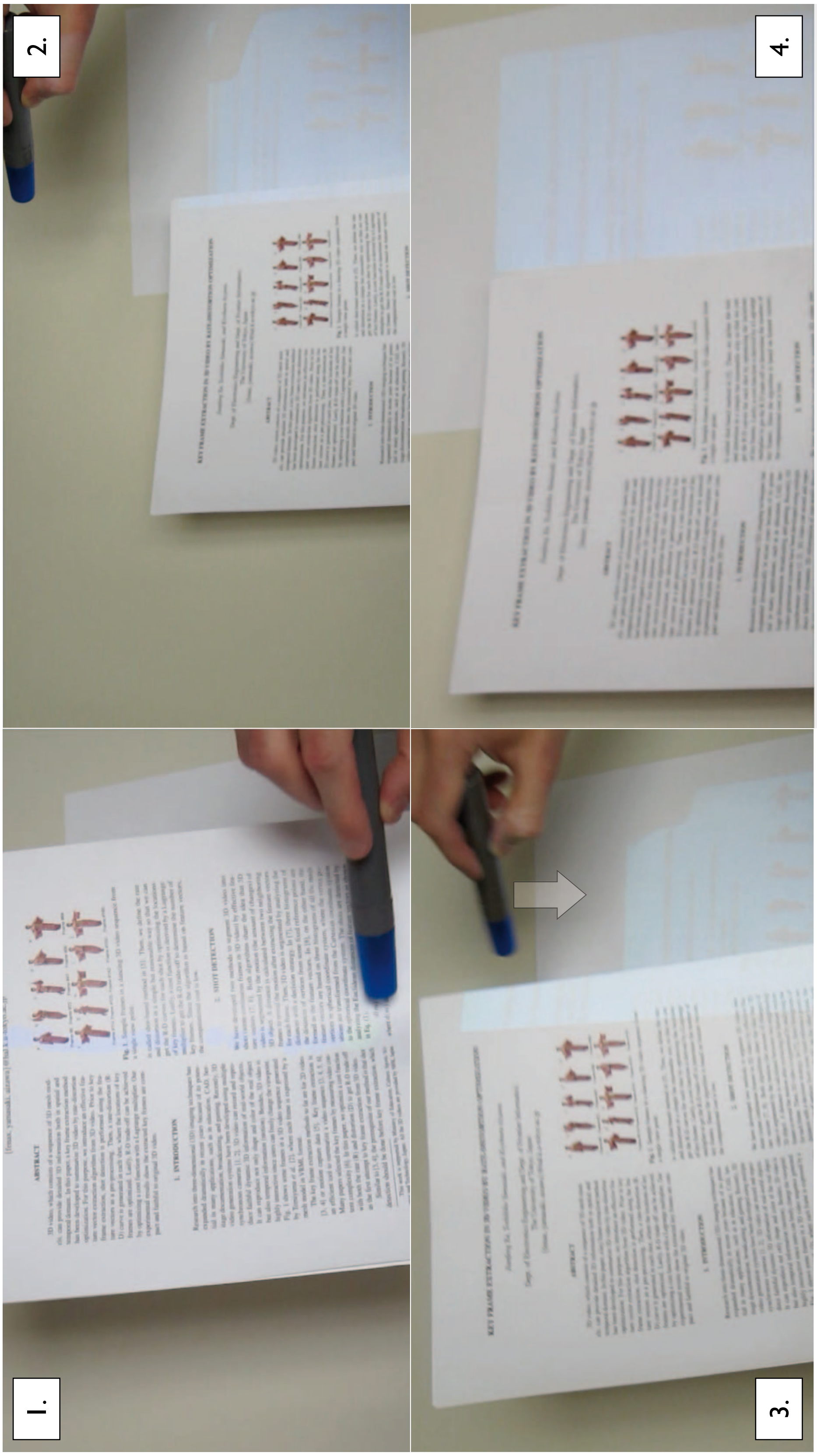


Figure 4.10: (1) The piece of paper is held in 3D space and a pen is used to select a part of the document (blue line). (2) The paragraph is in turn copied and projected into physical space. The pen can then be used to move the copied paragraph in 3D space and (3) paste it to surfaces in the vicinity by utilizing a simple flick gesture toward it (here: the paragraph is being pasted onto the table). (4) There, it can be for instance used for spatial comparison.

4.5 System Setup and Algorithms

We have prototypically implemented the interaction techniques. In the following, we describe the hardware setup, as well as the implemented algorithms.

4.5.1 Hardware

Figure 4.11 shows our prototype. We have attached an Aaxa L1 laser pico projector to a Microsoft Kinect with hook-and-loop tape, which we use as a mobile camera-projector unit. The projector has a resolution of 800 600 pixels. The Microsoft Kinect features a pair of depth-sensing range cameras (320 240 pixels), an infrared structured light source and a regular RGB color camera (640 480 pixels). In order to support hassle free document recognition, we have attached a megapixel webcam with *autofocus* to the unit. Kinect, webcam and pico projector are calibrated and aligned.

The mobile camera-projector unit can be further mounted onto a strong suction cup, which also features a handle. Thus the unit can be easily carried in one hand by using the handle. Moreover, it can be attached to basically any flat surface, even vertical surfaces or ceilings to achieve a top-down projection.

Figure 4.11: Hardware prototype using a Microsoft Kinect, mounted on a suction cup. The pico projector is placed on top of the Kinect. We have added a webcam on the right hand side for document recognition.

4.5.2 Algorithms

In the following, we describe the algorithms used to track objects, support the spatial interaction and recognize physical documents.

4.5.2.1 Object Tracking and Interaction Support

As projection surfaces, we currently consider flat surfaces of 3D objects. We model them as 2D planes in 3D space. To support a robust tracking of arbitrary objects, independent of varying lighting conditions, we aimed at using solely the depth image in our tracking algorithm. The algorithm is thus less complex than other approaches [Lowe, 2004], yet robust and highly efficient due to its simplicity. Algorithm 1 depicts a pseudocode representation of the algorithm.

First, a threshold is applied to the depth image to filter out any background objects (*line 2*). A blob detection for the objects in the scene is carried out (*line 3*). As a simple example, Figure 4.12 (left) shows only one object (here: a piece of paper), which is held in hand. Figure 4.12 (right) shows the corresponding depth image. We isolate the object from the scene (here: to discard the hand) in three steps, which are carried out for each detected blob (*line 4*):

1. **Breaking up weakly connected components:** the objective of this step is to detect weak connections between objects in the image and eliminate them to finally isolate the target object (i.e. the paper in Fig. 4.12). A weak connection is a thin line in the input image, connecting areas in the image which technically resemble one large blob (e.g. the piece of paper and the arm in Fig. 4.12). The separation is done with four basic image operations. First, an *and*-mask of the detected is applied to the image, to discard other blobs and therefore focus only on the current blob (*line 5*). The image is then blurred heavily, which results in lower gray-color values for the connections. Then a binary threshold is applied, eliminating the blurred borders (*line 6*). Finally, morphological open and close operators are applied to concretize the object borders (*line 7*).
2. **Detecting inner points of the target object:** the resulting image of step 1 contains isolated objects (i.e. both paper and hand in Fig. 4.12 are now two separate blobs). However, due to the image operations, the area and consequently the contour have been reduced. A further blob detection

now enables the detection of the reduced area (*line 8*). The algorithm choses the largest blob as the desired projection target (*line 9*).

3. **Mapping inner to original corner points:** a rotation invariant minimum bounding rectangle of the corresponding blob is calculated. The corner points of this bounding rectangle serve as the input points for the next step: the inner corner points are finally mapped to the original object corners by considering the contour of the object recognized in Figure 4.12. The bounding rectangle (and thus the inner corner points) is iteratively expanded by a fixed factor σ to approach the contour of the original target object (*lines 10-14*). Once the distance is smaller than a certain threshold δ, the corners of the target object have been found. The algorithm then stores the detected target object and starts over for the remaining blobs.

Algorithm 1 Object Tracking for LightBeam

```
1:  procedure TrackObjects(grayImage,objects)          ▷ objects serves as output set.
2:      threshold(grayImage)                            ▷ Apply depth threshold.
3:      blobs ← detectBlobs(grayImage)
4:      for each blob in blobs do
5:          img ← and(blob, grayImage)                  ▷ Remove other blobs.
6:          binaryThreshold(blurHeavily(img))
7:          dilate(erode(img))
8:          reducedBlobs ← detectBlobs(img)
9:          lBlob ← getLargestBlob(reducedBlobs)        ▷ Within original blob.
10:         contour ← getContour(blob)
11:         repeat
12:             targetObject ← lBlob
13:             lBlob ← expandArea(lBlob, σ)            ▷ Uniform expansion by factor σ.
14:             corners ← getCornerPoints(boundingRectangle(lBlob))
15:         until distance(contour, corners) ≤ δ        ▷ If near to original blob contour.
16:         objects.add(targetObject)
17:     end for
18:     return objects
19: end procedure
```

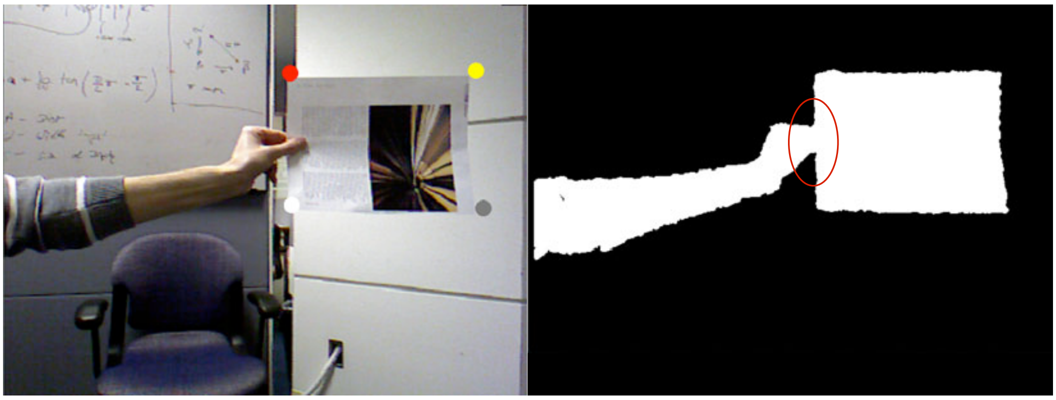


Figure 4.12: ***Left:*** **color image of a paper, held in hand. Its four corners are detected and indicated by four colored dots.** ***Right:*** **depth image after thresholding and blob detection. The red mark designates the thin connection, which the algorithm removes for object detection.**

In its current version, the algorithm is implemented using OpenCV[1]. It is important to note that line 9 in Alg. 1 restricts projection surfaces to be not smaller than a user's hand, assuming that objects in the scene are used for tangible interaction. The algorithm needs to be adapted to support smaller projection surfaces.

In combination with the depth information for the detected object contour, we model and track the detected objects as 2D planes in 3D space. The projection is mapped using a homography, correcting any perspective errors. We also analyze the optical flow within the regions of the blobs in the RGB image. This allows us to detect whether an object has been rotated. Additional interaction devices such as the pen in Figure 4.10 are tracked based on their color. As mentioned earlier, more sophisticated approaches such as touch have been described elsewhere [Harrison et al., 2011] and are out of the scope of this thesis.

4.5.2.2 Document Recognition

The system automatically recognizes paper documents to support the rich interactions described in the mobile document interaction scenario. The recognition uses FACT [Liao et al., 2010], which unitizes local natural features [Lowe,

[1]http://www.opencv.org (last checked October 29, 2012)

2004] to identify ordinary paper documents without any special markers. The current FACT implementation can operate at about 0.5 fps for recognizing a frame of 640×480 pixels on a PC with a quad core 2.8GHz CPU and 4GB RAM. Considering that users usually do not change documents very quickly during their tasks, this recognition speed is acceptable for practical use. The FACT implementation had to deal with various difficulties due to only using data from an RGB camera; e.g. small document tilting angles or interferences of overlaid projections with the original natural features.

FACT provides an interface which accepts captured camera images and returns the detected digital version of the document (i.e. a PDF). Figure 4.13 illustrates how the communication between LightBeam and FACT works. We leverage the capabilities of the Kinect depth camera to overcome these difficulties and enhance the camera image before passing it to FACT. The 3D pose estimation based on the depth image is independent of the document's natural features and thus the system is robust to insufficient feature correspondence. Moreover, a rectification of the color images based on the 3D pose decreases the perspective distortion and allows for greater tilting angles. Last, the pose estimation and the document recognition can be carried out in two separate threads, each updating the world model *asynchronously*. Therefore, from the aspect of users, the system is able to locate specific document content in 3D space in real time.

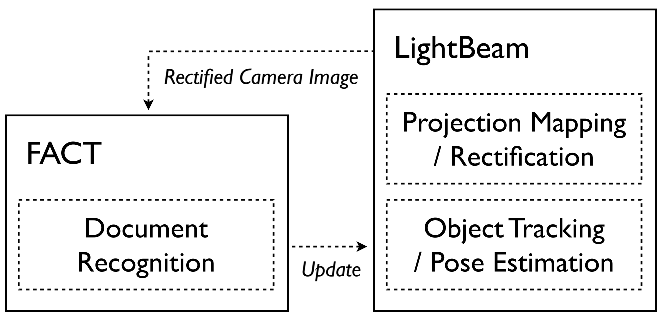


Figure 4.13: LightBeam separates the document recognition into two threads: it continuously estimates the 3D pose of a document and asynchronously queries FACT with rectified camera images. FACT then sends the recognized document back to LightBeam.

4.6 Early User Feedback

We have evaluated the prototypically implemented interaction techniques in an early user feedback session with interaction design researchers. Our main objective was to get a first impression whether the actual techniques are conceptually sound and how the experts would actually use them to interact with physical objects.

4.6.1 Setup and Methodology

We invited 6 interaction design researchers as a focus group to our living lab to evaluate the interaction techniques in semi-structured interviews. Our lab is an open space, containing desks (to simulate a working environment) and an area comparable to a living room with couches and a large LCD TV. The designers (all of them male) were in average 30 years of age and each of them had about 5 years of professional experience. Two of them also participated in our exploratory field study. The session lasted about 3 hours.

The interaction techniques were presented within the scope of each application scenario. The participants were asked to familiarize themselves with our hardware prototype. The desk contained typical items such as books, a laptop, pens, etc. They were given the opportunity to tryout each technique on their own, using the objects in the vicinity. Although our prototype requires to be wired to a PC for data transfer, the participants were able to roam around freely whilst carrying and repositioning the LightBeam. As data sources, we used the semi-structured interviews and also observed the participants. We transcribed the data and analyzed salient quotes.

4.6.2 Results and Discussion

All participants easily understood the interaction techniques. They liked the tight integration of physical objects and digital information, since "*this allows for a direct interaction with the virtual data*", as one participant noted.

The participants were focused primarily on the role of physical objects. Throughout the session, the participants repeatedly stressed the significance of using virtually any object to control the projection; in our example the rotation of objects. This also diminished their concerns that objects might lose their original function when being used as tangible controls. One participant commented: "*I like this kind of casual functional overlay. Now I am not afraid that I will end up with two co ee mugs on my table, since one might be dedicated to one speci c function*". However, they noted that they might want to bind certain types of information to special objects on purpose.

Moving any object into the beam to take a peek into the virtual world was considered important for supporting quick information access in-situ. It was considered particularly helpful when already dealing with physical objects, such as paper, on the table, since lifting them further into the beam triggered the seamless transition between different levels of detail. One participant commented: "*Projecting onto the table would be good, but actually, the table is too large, there is no frame*". The other participants agreed. This further underlines our findings from the exploratory study: physical objects provide natural frames.

When capturing physical objects within the beam, the participants again considered the casual overloading of physical objects (here: the pen) with digital functionality as useful. They reported that browsing and selecting digitally captured objects using the object movement in the z-direction is beneficial for providing an overview over and quick access to most recently captured objects. For larger collections however, two participants would have preferred to interact on the object itself, e.g. through a gesture-based interface instead of moving it through space.

4.7 Conclusion

This chapter contributed LightBeam, a novel model of interaction which leverages pico projectors as light beams', adding a new conceptual dimension to the pico projector design space. LightBeam provides a fundamentally different interaction space for tangible interaction than larger projection spaces. Being placed in a user's vicinity, it provides a dedicated interaction space through its highly limited projection ray. This tightly integrates both virtual and physical space and allows for both device- and tangible, space-centric interaction with real world objects–not just one particular mobile device.

In particular, this chapter made the following contributions:

- **Conceptual Framework:** We have contributed a detailed analysis of the conceptual framework of light beam with respect to prior pico projector research. Up to now, research on pico projector interaction emphasized on fixed and flat projection surfaces in physical space. The interaction was either projector-centered or relied on direct input on the projection surface. The tangible character of real world objects was not systematically explored for pico projector interaction before. Moreover, it was unclear how the mobility of physical objects could be actually leveraged for tangible interaction and what kind of projected information actually matched the affordances of physical objects.

- **Qualitative, Exploratory Field Study:** We investigated these points in a qualitative, explorative field study with 8 interaction design researchers, systematically explored the light beam concept and contributed our findings. We identified relevant theoretical dimensions comprising the projector placement and provision of both in- and output with real world objects. The results show that moving objects in the beam is a central theme for interaction in real world settings. Moving the projector–as predominantly advocated in previous research–is *not*. Objects provide a physical framing for projections and therefore embody them. Projections can be bound to objects of particular shape (e.g. boxes as large displays), but can be also adapted to deformable physical objects, depending on both application and user preferences.

- **Novel Interaction Techniques:** Based on the study results, we designed a set of interaction primitives and contributed several interaction techniques. The follow a central theme: moving objects into the beam charters them with both output and input functionality. Here, the highly limited projection ray plays an important role. It serves as a dedicated interaction hotspot wherein objects can be deliberately moved, therefore overloading the objects' original mapping (e.g. using a cup as a tangible control instead of drinking from it). With-drawing the objects from the beam then removes the over-loaded and respectively restores the original mapping. For instance by leveraging physical affordances of objects for tangible controls instead of dedicating specific objects to specific functions, we provide a loose coupling between object and functionality. This is key for object-based interactions in nomadic settings, where it cannot be taken for granted that specific objects are available.

- **Evaluation:** The implemented interaction techniques were evaluated in an early user feedback session with interaction design researchers. The results confirm the identified theoretical dimensions from the first study and underline the importance of the tight integration of both virtual and physical space. The interviews indicate that the implemented techniques particularly support users when carrying out tasks such as mobile document interaction. Furthermore, the results underline the importance of direct interaction and physical mappings.

The results also indicate a tension between device- and space-centric interaction. Externalizing the virtual information space to the physical space and providing space-centric interaction support is certainly helpful, e.g. for cognitive offloading when (re-)structuring information and laying it out in physical space. However, physical objects also become tangible interaction devices. Thus users can adopt a device-centric interaction perspective to browse an information space using a particular object, e.g. through gesture-based or tilt-based interaction.

The latter observation leads to novel, open research questions, such as when and how seamlessly should users be able to switch between the two interaction perspectives. Furthermore, with the current trend toward multiple mobile devices per user, future research should also consider the interplay of multiple objects as devices and using multiple light beams for nomadic pico projector interaction with real world objects.

We believe that object-based interactions, in combination with the casual overloading of physical mappings and already existing touch-based interfaces [Harrison et al., 2011], will fundamentally change how we ubiquitously interact with augmented real-world objects in nomadic settings.

Chapter 5
Conclusions

Increasingly capable mobile devices such as the Apple iPhone shape the way we engage with multimedia information. Being foremost sophisticated information devices, they provide the ubiquitous availability of a variety of multimedia information while being on the move; ready at our fingertips, virtually *anywhere* and *anytime*. At the same time, these devices generate a need for efficient information exploration and rendering support. While the latter is being addressed by industry with recent advancements in high definition displays, interfaces for mobile multimedia information exploration are in their infancies. Traditional user interface concepts cannot be transferred as-is and with novel interaction paradigms, such as touch or physical input, on the horizon, it is unclear how to design for the mobile interaction with large multimedia information spaces.

The overarching goal of this thesis was to explore novel, more usable and enjoyable ways for interacting with large multimedia information spaces on mobile devices. This thesis followed three main research directions:

1. *Device-Centric Interaction*, leveraging the very affordances of mobile devices for efficient user interfaces albeit the restrictive form factors.
2. *Space-Centric Interaction*, pushing the virtual information space across the boundaries of the small screen toward the physical space.
3. *Integrating Device- and Space-Centric Interaction*, which investigates how virtual and physical space can be tightly integrated to foster interaction with real world objects–not just one particular mobile device.

All research directions followed an empirically-inductive research methodology to develop novel user interfaces and interaction designs, which in turn have undergone a user-centered design process.

In this final chapter, we revisit the contributions of each research direction, summarize the main outcomes of this thesis and last, point out directions of future research.

5.1 Summary

In the following, we summarize the main outcomes and contributions of this thesis according to each research direction.

Device-Centric Interaction. In this research direction, we adopted a *broad view on the design space* of mobile interaction with topically-interrelated video collections, being a prime example for the mobile interaction with large multimedia information spaces. Informed by an *analysis of use patterns of mobile video browsing* and *participatory design sessions*, we set up a design space that covers two dimensions: the broad interaction metaphor used in the interaction concept (GUI-based, gesture-based, physical) and the complexity of the navigation. This enabled us to *systematically derive 8 interaction concepts* (7 novel concepts, one standard interface), which are situated within the design space and implemented on the iPhone.

As part of the user-centered design process, we conducted a *controlled experiment with 44 participants* and collected and analyzed more than *18 hours of video observations*. The results provide *empirical* evidence that designers should leverage the novel capabilities of mobile devices, such as direct touch and inertial sensors. A more traditional GUI approach, as in this case the iPhone video player, is likely to lead to lower efficiency and is more error-prone. The usability error analysis shows that even a simple misplacement of interface elements can lead to the loss of internal locus of control and therefore to severe usability breakdowns. Moreover, the error analysis underlines the potential of gesture-based or physical interfaces for mobile video browsing.

Our analysis also provided the basis for *design principles* for mobile video browsers. By supporting spatiotemporal browsing metaphors and discrete temporal navigation and by placing interface elements carefully, designers can improve both usability and user experience of future mobile video browsers.

Space-Centric Interaction. In this research direction, we extended the virtual information space to the physical space. For this purpose, we utilized mobile devices as *spatially aware see-through displays*, providing an embodied dynamic peephole onto the large virtual information space. The analysis of prior work in this field showed that most of the approaches were evaluated in lab settings and not in a real world context, where they actually should be applied. Thus, the field lacked a *fundamental understanding of how users would actually interact through embodied peepholes in mobile, real world settings*. To fill this void, we conducted a qualitative, exploratory field study. This allowed us to derive an *empirically grounded theory* of embodied peephole interaction.

Based upon the developed theory, we derived *design implications* for future embodied peephole interfaces, which in turn led to the *prototypical implementation of novel interaction techniques*. The theory also revealed that the layout of the information space in physical space is highly important: poorly positioned content might not be reachable. The overcome this problem, the layout can be optimized by modeling a user's performance.

This observation lead to the contribution of a *novel movement time model for the embodied navigation of a-priori unknown information spaces* with spatially-aware displays. The model is inspired by physiological aspects of the human body and thus particularly addresses the affordances of embodied interaction with such displays. We contributed the results of a *controlled experiment* with 32 participants using a physical apparatus validate the mathematical model. The results provide empirical evidence that it satisfactorily models the movement time in a-priori known spaces–as good as existing approaches. However, in the case of an unfamiliar information space, the results provide significant evidence that the search time is better modeled by our novel formulae.

Integrating Device- and Space-Centric Interaction. We explored using pico projectors as light beams, bridging both of the aforementioned interaction perspectives. We have contributed a *detailed analysis of the conceptual framework* of light beam with respect to prior pico projector research. Up to now, research on pico projector interaction emphasized on fixed and flat projection surfaces in physical space. The interaction was either projector-centered or relied on direct input on the projection surface. The *tangible character of real world objects* was not systematically explored for pico projector interaction before. The overarching goal of this research direction was to fill this void.

As part of the qualitative exploration, we conducted an *exploratory eld study* to get a deeper understanding of how the concept of light beam can be applied in *real world scenarios*. We systematically and iteratively explored the concept with 8 interaction design researchers in different usage scenarios, such as a café or their working environment. We identified relevant theoretical dimensions comprising the projector placement and how to provide both in- and output with real world objects.

This allowed us to advance the field of pico projector interaction by designing *novel interaction techniques* which go beyond prior work. The implemented interaction techniques were evaluated in *expert interviews* with interaction design researchers. The results confirmed the identified theoretical dimensions from the first study and underline the importance of the tight integration of both virtual and physical space. The results also indicated a tension between device- and space-centric interaction. Externalizing the virtual information space to the physical space and providing space-centric interaction support is certainly helpful, e.g. for cognitive offloading when (re-)structuring information and laying it out in physical space. However, physical objects also become tangible interaction devices. Thus users can adopt a device-centric interaction perspective to browse an information space using a particular object in-situ, e.g. through gesture-based or tilt-based interaction.

5.2 Future Research Directions

Each of the three research directions followed in this thesis yield immediate, open research questions, which should be considered for future research.

In terms of *device-centric interaction*, future research needs to be carried out to fully grasp and understand the potential of physical interfaces (e.g. tilt-based techniques). Moreover, long-term studies are needed to explore the usefulness of the presented interfaces for promising application domains such as mobile, technology-enhanced learning.

With respect to *space-centric interaction*, two research directions seem apparent. On the one hand, the techniques developed for embodied dynamic peephole interaction mainly focused on the *exploration* of existing multimedia information spaces. Future research shall focus on the creation and re-structuring of virtual, spatially-aware information spaces in mobile settings. On the other hand, the movement time model for embodied dynamic peephole pointing focused on one-dimensional pointing as a first step towards multi-dimensional models. Future research shall investigate possible extensions toward two- and three-dimensional models, eventually.

Future research on *integrating device- and space-centric interaction* shall explore novel input modalities for LightBeam, e.g. by combining both touch and tangible interaction. Moreover, with the current trend toward multiple mobile devices per user, future research shall also consider the interplay of multiple objects as devices and using multiple light beams for nomadic pico projector interaction with real world objects.

Mobile interaction itself is a highly dynamic and vibrant field of research, operating at a high pace. Crucial to its advancement is technology support. But with the advent of enabling technologies such as the Kinect, prototyping novel interface concepts becomes feasible. Furthermore, technology-centered conferences such as the yearly ACM Symposium on User Interface Software and Technology contribute equally. In addition to the detailed aspects for future work mentioned above, we enumerate a few broader technology-centered research directions in the following.

Flexible, Paper-like Displays. With recent advancements in display technology, flexible, paper-like displays are not far from production. Samsung and LG both showed off physical, working prototypes at several technology fairs like CES in 2011 and 2012. This paves the way for novel interaction concepts, inspired by what is used in everyday practices: analogue, physical paper. Consequently, new interfaces can benefit from the flexibility of paper, dynamically increasing or decreasing display space–*on demand*. One example is Xpaaand [Khalilbeigi et al., 2011], a projection-based prototype of a rollable display, investigating rolling as a core interaction technique for future digital paper-like displays. Dynamic folding of paper-like displays has been investigated by Khalilbeigi et al. [2012]. With respect to the mobile interaction in large multimedia information spaces, one aspect becomes apparent: the information space on "paper" can be laid out in physical space, therefore allowing for spatial interaction. While Girouard et al. [2012] have investigated stacking of such displays and Lissermann et al. [2012] developed spatial techniques for video interaction, we feel that this is a promising research direction for mobile interaction with large multimedia information spaces and can be seen analogously to the work with physical paper documents. One obvious and yet unexplored research question is how and if this analogon can be leveraged to design even more usable interface concepts for large information spaces.

Immersive Mixed Reality. Parts of this thesis focused on bridging the digital-physical-divide through mixed and augmented reality. While the concept of LightBeam (as presented in chapter 4), provides a compelling and tight coupling between physical and digital artifacts, it is still projection-based. Apart from the trend of flexible, paper-like displays, wearable electronics suggest to aid in creating an even more immersive coupling. A recent and promising advancement in this field is Google's project "Glass": a pair of multi-modal augmented reality goggles which display the user interface directly in user's field of view. They feature multi-modal in- and output. In terms of interacting with large multimedia information spaces, this would enable a more intriguing mapping of the digital content to the physical space. Combined with haptic feedback (cf. Iwamoto et al. [2008]), this would allow for a truly immersive and foremost mobile exploration of large multimedia information spaces.

Designing Across Devices. The results of this thesis show how important it is to design for a user's mental model, particularly when interacting with large multimedia information spaces under constraints, such as a small display. This is not a surprising fact in general, given that Donald Norman postulated the importance of a shared common model of both system and user in his seminal work on the design of everyday things [Norman, 2002]. However, we currently face a trend in that users tend to own not a *single* personal device, but a *manifold set of personal devices*: a mobile phone, a tablet, an ebook reader, a laptop or a desktop computer. Ubiquitous computing does not stop here: devices such as in-car entertainment systems or shared public displays allow users to consume multimedia content virtually anywhere they want. They interact with them throughout the day, sequentially, not necessarily at the same time. However, the interaction models on and across these devices differ greatly and in turn, they do not match the user's mental model generically. As a consequence, we propose to not only concentrate on interacting with large multimedia information spaces on a single device, but *supporting a user's mental model across devices*. We speculate that this will greatly improve the user experience: users shall be able to interact with content without caring about the actual device.

Bibliography

John Adcock, Matthew Cooper, Laurent Denoue, Hamed Pirsiavash, and Lawrence A. Rowe. Talkminer: a lecture webcast search engine. In *Proceedings of the international conference on Multimedia*, MM '10, pages 241–250, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-933-6. doi: 10.1145/1873951.1873986. URL `http://doi.acm.org/10.1145/1873951.1873986`. (Cited on page 23.)

Tue Haste Andersen. A simple movement time model for scrolling. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1180–1183, New York, NY, USA, 2005. ACM. (Cited on pages 100 and 101.)

Patrick Baudisch and Ruth Rosenholtz. Halo: a technique for visualizing off-screen objects. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, page 481, New York, New York, USA, 2003. ACM Press. ISBN 1581136307. doi: 10.1145/642611.642695. URL `http://portal.acm.org/citation.cfm?doid=642611.642695`. (Cited on pages 82 and 87.)

DV Beard and JQ Walker. Navigational techniques to improve the display of large 2-dimensional spaces. *BEHAVIOUR & INFORMATION TECHNOLOGY*, 9 (6):451–466, NOV-DEC 1990. ISSN 0144-929X. (Cited on page 2.)

Mark Billinghurst, H Kato, and I Poupyrev. The MagicBook—moving seamlessly between reality and virtuality. *IEEE Computer Graphics and Applications*, 2001. URL `http://doi.ieeecomputersociety.org/10.110910.1109/38.920621`. (Cited on page 78.)

Richard A. Bolt. "Put-that-there": Voice and gesture at the graphics interface. *ACM SIGGRAPH Computer Graphics*, 14(3):262–270, July 1980. ISSN 00978930. doi: 10.1145/965105.807503. URL `http://dl.acm.org/citation.cfm?id=965105.807503`. (Cited on page 138.)

Sebastian Boring, Sven Gehring, Alexander Wiethoff, Anna Magdalena Blöckner, Johannes Schöning, and Andreas Butz. Multi-user interaction on media facades through live video on mobile devices. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2721–2724, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979342. URL http://doi.acm.org/10.1145/1978942.1979342. (Cited on page 80.)

John Brooke. SUS: a "quick and dirty" usability scale. *Usability Evaluation in Industry*, 1996. (Cited on page 53.)

Jason A. Brotherton and Gregory D. Abowd. Lessons learned from eclass: Assessing automated capture and access in the classroom. *ACM Trans. Comput.-Hum. Interact.*, 11(2):121–155, June 2004. ISSN 1073-0516. doi: 10.1145/1005361.1005362. URL http://doi.acm.org/10.1145/1005361.1005362. (Cited on page 22.)

Andreas Butz and Antonio Krüger. A generalized peephole metaphor for augmented reality and instrumented environments. In *Workshop on Software Technology for Augmented Reality Systems*, 2003. (Cited on page 80.)

Andreas Butz and Antonio Krüger. Applying the peephole metaphor in a mixed-reality room. *IEEE Comput. Graph. Appl.*, 26(1):56–63, January 2006. ISSN 0272-1716. doi: 10.1109/MCG.2006.10. URL http://dx.doi.org/10.1109/MCG.2006.10. (Cited on page 80.)

Xiang Cao and Ravin Balakrishnan. Interacting with dynamically defined information spaces using a handheld projector and a pen. In *Symposium on User Interface Software and Technology*, 2006. URL http://portal.acm.org/citation.cfm?id=1166253.1166289. (Cited on pages 80 and 122.)

Xiang Cao, Clifton Forlines, and Ravin Balakrishnan. Multi-user interaction using handheld projectors. In *Symposium on User Interface Software and Technology*, 2007. URL http://portal.acm.org/citation.cfm?id=1294211.1294220. (Cited on page 122.)

Xiang Cao, Jacky Jie Li, and Ravin Balakrishnan. Peephole pointing: modeling acquisition of dynamically revealed targets. In *CHI '08: Proceeding of* *the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1699–1708, New York, NY, USA, 2008. ACM. (Cited on pages 99, 100, 103 and 113.)

Jessica R. Cauchard, Mike Fraser, Teng Han, and Sriram Subramanian. Steerable projection: exploring alignment in interactive mobile displays. *Personal Ubiquitous Comput.*, 16(1):27–37, January 2012. ISSN 1617-4909. doi: 10.1007/s00779-011-0375-3. URL http://dx.doi.org/10.1007/s00779-011-0375-3. (Cited on page 121.)

Kai-Yin Cheng, Sheng-Jie Luo, Bing-Yu Chen, and Hao-Hua Chu. Smartplayer: user-centric video fast-forwarding. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 789–798, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1518823. URL http://doi.acm.org/10.1145/1518701.1518823. (Cited on page 21.)

Kai-Yin Cheng, Rong-Hao Liang, Bing-Yu Chen, Rung-Huei Laing, and Sy-Yen Kuo. iCon: utilizing everyday objects as additional, auxiliary and instant tabletop controllers. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, pages 1155–1164, New York, New York, USA, April 2010. ACM Press. ISBN 9781605589299. doi: 10.1145/1753326.1753499. URL http://dl.acm.org/citation.cfm?id=1753326.1753499. (Cited on page 136.)

Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys*, 41(1):1–31, 2008. ISSN 03600300. (Cited on pages 82 and 98.)

Lisa G. Cowan and Kevin A. Li. Shadowpuppets: supporting collocated interaction with mobile projector phones using hand shadows. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2707–2716, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979340. URL http://doi.acm.org/10.1145/1978942.1979340. (Cited on page 121.)

Lisa G. Cowan, Nadir Weibel, William G. Griswold, Laura R. Pina, and James D.

Hollan. Projector phone use: practices and social implications. *Personal Ubiquitous Comput.*, 16(1):53–63, January 2012. ISSN 1617-4909. doi: 10.1007/s00779-011-0377-1. URL http://dx.doi.org/10.1007/s00779-011-0377-1. (Cited on page 119.)

Raimund Dachselt and Robert Buchholz. Natural throw and tilt interaction between mobile phones and distant displays. In *Conference on Human Factors in Computing Systems*, 2009. URL http://portal.acm.org/citation.cfm?id=1520340.1520467. (Cited on page 27.)

Raimund Dachselt, Jonna Häkkilä, Matt Jones, Markus Löchtefeld, Michael Rohs, and Enrico Rukzio. Pico projectors: firefly or bright future? *interactions*, 19(2):24–29, March 2012. ISSN 1072-5520. doi: 10.1145/2090150.2090158. URL http://doi.acm.org/10.1145/2090150.2090158. (Cited on page 3.)

Jim Dalrymple. Apple: iTunes u tops 600 million downloads (last checked: October 29, 2012), 09 2011. URL http://www.loopinsight.com/2011/09/07/itunes-u-educating-the-world/. (Cited on page 1.)

Manfred Del Fabro, Klaus Schoeffmann, and Laszlo Böszörmenyi. Instant video browsing: a tool for fast non-sequential hierarchical video browsing. In *Proceedings of the 6th international conference on HCI in work and learning, life and leisure: workgroup human-computer interaction and usability engineering*, USAB'10, pages 443–446, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-16606-7, 978-3-642-16606-8. URL http://dl.acm.org/citation.cfm?id=1947789.1947827. (Cited on page 24.)

Pierre Dragicevic, Gonzalo Ramos, Jacobo Bibliowitcz, Derek Nowrouzezahrai, Ravin Balakrishnan, and Karan Singh. Video browsing by direct manipulation. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 237–246, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357096. URL http://doi.acm.org/10.1145/1357054.1357096. (Cited on pages 20, 21 and 179.)

Heiko Drewes. Only one fitts' law formula please! In *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors* *in computing systems*, pages 2813–2822, New York, NY, USA, 2010. ACM. (Cited on page 100.)

Deborah M. Edwards and Lynda Hardman. *Lost in hyperspace: cognitive mapping and navigation in a hypertext environment*, pages 90–105. Hypertext: theory into practice. Intellect Books, Exeter, UK, 1999. ISBN 1-871516-28-5. (Cited on pages 49 and 82.)

Ben Falchuk, Alex Glasman, and Konstantin Glasman. A tool for video content understanding on mobile smartphones. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, MobileHCI '08, pages 307–310, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-952-4. doi: http://doi.acm.org/10.1145/1409240.1409275. (Cited on page 26.)

P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. 1954. *J Exp Psychol Gen*, 121(3):262–269, September 1992. (Cited on page 100.)

George W. Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, 36(7):39–49, 1993. ISSN 00010782. doi: 10.1145/159544.159566. URL `http://portal.acm.org/citation.cfm?doid=159544.159566`. (Cited on pages 3 and 80.)

Audrey Girouard, Aneesh Tarun, and Roel Vertegaal. Displaystacks: interaction techniques for stacks of flexible thin-film displays. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 2431–2440, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2208276.2208406. URL `http://doi.acm.org/10.1145/2208276.2208406`. (Cited on page 155.)

B. G. Glaser and A. Strauss. The discovery of grounded theory: Strategies for qualitative research. 1967. (Cited on page 11.)

James Glass, Timothy J. Hazen, Scott Cyphers, Igor Malioutov, David Huynh, and Regina Barzilay. Recent Progress in the MIT Spoken Lecture Processing Project. In *Proc. Interspeech*, 2007. (Cited on pages 23 and 179.)

James R. Glass, Timothy J. Hazen, D. Scott Cyphers, Ken Schutte, and Alex Park. The mit spoken lecture processing project. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, HLT-Demo '05, pages 28–29, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics. doi: http://dx.doi.org/10.3115/1225733.1225748. (Cited on page 23.)

Jun Gong and Peter Tarasewich. Guidelines for handheld mobile device interface design. In *Proceedings of Designing Interactive Systems '04*, 2004. doi: 10.1.1.87.5230. (Cited on pages 2 and 17.)

Sean Gustafson, Patrick Baudisch, Carl Gutwin, and Pourang Irani. Wedge: clutter-free visualization of off-screen locations. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 787, New York, New York, USA, 2008. ACM Press. ISBN 9781605580111. doi: 10.1145/1357054.1357179. URL `http://portal.acm.org/citation.cfm?doid=1357054.1357179`. (Cited on page 82.)

Chris Harrison, Desney Tan, and Dan Morris. Skinput: appropriating the body as an input surface. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, pages 453–462, New York, New York, USA, April 2010. ACM Press. ISBN 9781605589299. doi: 10.1145/1753326.1753394. URL `http://dl.acm.org/citation.cfm?id=1753326.1753394`. (Cited on page 122.)

Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 441–450, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047255. URL `http://doi.acm.org/10.1145/2047196.2047255`. (Cited on pages 123, 139, 145 and 150.)

Marc Hassenzahl, Axel Platz, Michael Burmester, and Katrin Lehner. Hedonic and ergonomic quality aspects determine a software's appeal. In *Proceedings of the 2000 annual conference on Human factors in computing systems - CHI '00*, pages 201–208. ACM New York, NY, USA, 2000. (Cited on page 53.)

Marc Hassenzahl, Michael Burmester, and Franz Koller. Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität.

In J. Ziegler and G. Szwillus, editors, *Mensch & Computer 2003. Interaktion in Bewegung*, pages 187–196. B.G. Teubner, 2003. (Cited on pages 59 and 60.)

Alexander Haubold, Promiti Dutta, and John R. Kender. Evaluation of video browser features and user interaction with VAST MM. In *Proceedings of ACM Multimedia 2008*. ACM New York, NY, USA, 2008. URL `http://portal.acm.org/citation.cfm?doid=1459359.1459419`. (Cited on page 24.)

A. Hauptmann. Lessons for the future from a decade of informedia video analysis research. *Image and Video Retrieval*, pages 595–595, 2005. (Cited on page 24.)

Niels Henze and Susanne Boll. Evaluation of an off-screen visualization for magic lens and dynamic peephole interfaces. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 191–194, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-835-3. doi: 10.1145/1851600.1851632. URL `http://doi.acm.org/10.1145/1851600.1851632`. (Cited on page 82.)

Benjamin Höferlin, Markus Höferlin, Daniel Weiskopf, and Gunther Heidemann. Information-based adaptive fast-forward for visual surveillance. *Multimedia Tools Appl.*, 55(1):127–150, October 2011. ISSN 1380-7501. doi: 10.1007/s11042-010-0606-z. URL `http://dx.doi.org/10.1007/s11042-010-0606-z`. (Cited on page 22.)

David Holman, Roel Vertegaal, Mark Altosaar, Nikolaus Troje, and Derek Johns. Paper windows: interaction techniques for digital paper. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, CHI '05, pages 591–599, New York, New York, USA, 2005. ACM Press. ISBN 1581139985. doi: 10.1145/1054972.1055054. URL `http://doi.acm.org/10.1145/1054972.1055054`. (Cited on page 123.)

Kasper Hornbaek, Benjamin B. Bederson, and Catherine Plaisant. Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Trans. Comput.-Hum. Interact.*, 9(4):362–389, December 2002. ISSN 1073-0516. doi: 10.1145/586081.586086. URL `http://doi.acm.org/10.1145/586081.586086`. (Cited on page 2.)

Jochen Huber. Exploring the Mobile Interaction with Large Information Spaces within Mixed Reality Environments: A Grounded Theory Approach. In *Begreifbare Interaktionen in gemischten Wirklichkeiten, Workshop Proceedings of Mensch & Computer 2010*, pages 151–156. GI LNI, September 2010. (Cited on page 13.)

Jochen Huber, Jürgen Steimle, Roman Lissermann, Simon Olberding, and Max Mühlhäuser. Wipe'n'Watch: Spatial Interaction Techniques for Interrelated Video Collections on Mobile Devices. In *BCS HCI '10: Proceedings of the 2010 British Computer Society Conference on Human-Computer Interaction*, pages 423–437. British Computer Society (BCS), September 2010a. (Cited on page 13.)

Jochen Huber, Jürgen Steimle, and Max Mühlhäuser. Mobile interaction techniques for interrelated videos. In *CHI '10 extended abstracts on Human factors in computing systems*, pages 3535–3540. ACM New York, NY, USA, April 2010b. (Cited on page 13.)

Jochen Huber, Jürgen Steimle, and Max Mühlhäuser. Toward More Efficient User Interfaces for Mobile Video Browsing. In *Proceedings of the international conference on Multimedia - MM '10*, MM '10, pages 341–350, New York, New York, USA, October 2010c. ACM Press. ISBN 9781605589336. doi: 10.1145/1873951.1873999. (Cited on page 13.)

Jochen Huber, Jürgen Steimle, and Max Mühlhäuser. Interaction Techniques for Mobile E-Lectures. In *Workshop on "Next Generation of HCI and Education" in conjunction with CHI 2010*, New York, NY, USA, April 2010d. ACM. (Cited on page 13.)

Jochen Huber, Jürgen Steimle, Simon Olberding, Roman Lissermann, and Max Mühlhäuser. Browsing E-Lecture Libraries on Mobile Devices: A Spatial Interaction Concept. In *2010 10th IEEE International Conference on Advanced Learning Technologies*, pages 151–155. IEEE, July 2010e. (Cited on page 13.)

Jochen Huber, Chunyuan Liao, Jürgen Steimle, and Qiong Liu. Toward Bimanual Interactions with Mobile Projectors on Arbitrary Surfaces. In *Proceedings of MP^2: Workshop on Mobile and Personal Projection in conjunction with CHI 2011*. CERT, May 2011a. (Cited on page 13.)

Jochen Huber, Jürgen Steimle, and Max Mühlhäuser. A Model of Embodied Dynamic Peephole Pointing for Hidden Targets. In *BCS HCI '11: Proceedings of the 2011 British Computer Society Conference on Human-Computer Interaction*, pages 315–320. British Computer Society (BCS), September 2011b. (Cited on page 13.)

Jochen Huber, Jürgen Steimle, Chunyuan Liao, Qiong Liu, and Max Mühlhäuser. LightBeam: Nomadic Pico Projector Interaction with Real World Objects. In *CHI '12: Extended Abstracts of the 30th International Conference on Human Factors in Computing Systems*, pages 2513–2518. ACM New York, NY, USA, May 2012a. (Cited on page 13.)

Jochen Huber, Jürgen Steimle, Chunyuan Liao, Qiong Liu, and Max Mühlhäuser. Lightbeam: Interacting with augmented real-world objects in pico projections. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM '12)*, page (to appear). ACM New York, NY, USA, December 2012b. (Cited on page 13.)

Wolfgang Hürst. Interactive audio-visual video browsing. In *Proceedings of the 14th annual ACM international conference on Multimedia*, MULTIMEDIA '06, pages 675–678, New York, NY, USA, 2006. ACM. (Cited on page 25.)

Wolfgang Hürst and Georg Götz. Interface issues for interactive navigation and browsing of recorded lectures and presentations. In *Proceedings of ED-MEDIA 2004, World Conference on Educational Multimedia, Hypermedia & Telecommunications*, pages 4464–4469. AACE Press, 2004. (Cited on page 22.)

Wolfgang Hürst and Konrad Meier. Interfaces for timeline-based mobile video browsing. *International Multimedia Conference*, 2008. URL `http://portal.acm.org/citation.cfm?doid=1459359.1459422`. (Cited on page 25.)

Wolfgang Hürst and Philipp Merkle. One-handed mobile video browsing. In *Proceedings of the 1st international conference on Designing interactive user experiences for TV and video*, UXTV '08, pages 169–178, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-100-2. doi: 10.1145/1453805.1453839. URL `http://doi.acm.org/10.1145/1453805.1453839`. (Cited on page 26.)

Wolfgang Hürst, Georg Götz, and Philipp Jarvers. Advanced user interfaces for dynamic video browsing. In *Proceedings of the 12th annual ACM international conference on Multimedia*, MULTIMEDIA '04, pages 742–743, New York, NY, USA, 2004. ACM. ISBN 1-58113-893-8. doi: 10.1145/1027527.1027694. URL http://doi.acm.org/10.1145/1027527.1027694. (Cited on page 21.)

Wolfgang Hürst, Georg Götz, and Martina Welte. Interactive video browsing on mobile devices. *International Multimedia Conference*, 2007a. URL http://portal.acm.org/citation.cfm?doid=1291233.1291284. (Cited on page 25.)

Wolfgang Hürst, Martina Welte, and Sabine Jung. An evaluation of the mobile usage of e-lecture podcasts. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, Mobility '07, pages 16–23, New York, NY, USA, 2007b. ACM. ISBN 978-1-59593-819-0. doi: http://doi.acm.org/10.1145/1378063.1378067. URL http://doi.acm.org/10.1145/1378063.1378067. (Cited on pages 16 and 18.)

Wolfgang Hürst, Konrad Meier, and Georg Götz. Timeline-based video browsing on handheld devices. *International Multimedia Conference*, 2008. URL http://portal.acm.org/citation.cfm?doid=1459359.1459545. (Cited on page 26.)

iTunes U. Overview webpage (last checked: October 29, 2012), 2012. URL http://www.apple.com/education/itunes-u/. (Cited on page 1.)

Takayuki Iwamoto, Mari Tatezono, Takayuki Hoshi, and Hiroyuki Shinoda. Airborne ultrasound tactile display. In *ACM SIGGRAPH 2008 new tech demos*, SIGGRAPH '08, pages 1:1–1:1, New York, NY, USA, 2008. ACM. doi: 10.1145/1401615.1401616. URL http://doi.acm.org/10.1145/1401615.1401616. (Cited on page 156.)

Matt Jones and Gary Marsden. *Mobile Interaction Design*. John Wiley & Sons, February 2006. ISBN 0470090898. URL http://www.amazon.com/exec/

obidos/redirect?tag=citeulike07-20&path=ASIN/0470090898. (Cited on pages 1 and 2.)

Susanne Jul and George W. Furnas. Critical zones in desert fog: aids to multiscale navigation. In *Proc. UIST '98*, pages 97–106, New York, NY, USA, 1998. ACM. ISBN 1-58113-034-1. doi: http://doi.acm.org/10.1145/288392.288578. (Cited on pages 2 and 82.)

Maryam Kamvar, Patrick Chiu, Lynn Wilcox, Sandeep Casi, and Surapong Lertsithichai. Minimedia surfer: browsing video segments on small displays. In *CHI '04 extended abstracts on Human factors in computing systems*, CHI EA '04, pages 1371–1374, New York, NY, USA, 2004. ACM. ISBN 1-58113-703-6. doi: http://doi.acm.org/10.1145/985921.986067. URL http://doi.acm.org/10.1145/985921.986067. (Cited on page 26.)

Shaun K. Kane, Daniel Avrahami, Jacob O. Wobbrock, Beverly Harrison, Adam D. Rea, Matthai Philipose, and Anthony LaMarca. Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In *Symposium on User Interface Software and Technology*, 2009. URL http://portal.acm.org/citation.cfm?id=1622176.1622202. (Cited on page 121.)

Toshikazu Karitsuka and Kosuke Sato. A Wearable Mixed Reality with an On-Board Projector. In *ISMAR '03 Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 321–322, October 2003. ISBN 0-7695-2006-5. URL http://dl.acm.org/citation.cfm?id=946248.946820. (Cited on page 122.)

Thorsten Karrer, Malte Weiss, Eric Lee, and Jan Borchers. Dragon: a direct manipulation interface for frame-accurate in-scene video navigation. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 247–250, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: http://doi.acm.org/10.1145/1357054.1357097. (Cited on pages 20 and 26.)

Thorsten Karrer, Moritz Wittenhagen, and Jan Borchers. Pocketdragon: a direct manipulation video navigation interface for mobile devices. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 47:1–47:3, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-281-8. doi: http://doi.acm.org/10.1145/1613858.1613917. URL http://doi.acm.org/10.1145/1613858.1613917. (Cited on pages 17 and 26.)

Mohammadreza Khalilbeigi, Roman Lissermann, Max Mühlhäuser, and Jürgen Steimle. Xpaaand: interaction techniques for rollable displays. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2729–2732, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979344. URL http://doi.acm.org/10.1145/1978942.1979344. (Cited on page 155.)

Mohammadreza Khalilbeigi, Roman Lissermann, Wolfgang Kleine, and Jürgen Steimle. Foldme: interacting with double-sided foldable displays. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 33–40, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1174-8. doi: 10.1145/2148131.2148142. URL http://doi.acm.org/10.1145/2148131.2148142. (Cited on page 155.)

Georg Klein and David Murray. Parallel tracking and mapping on a camera phone. In *Proc. ISMAR '09*, Orlando, October 2009. (Cited on page 78.)

Lecturnity. (last checked: October 29, 2012), 2012. URL http://www.lecturnity.com. (Cited on page 22.)

Chunyuan Liao, Hao Tang, Qiong Liu, Patrick Chiu, and Francine Chen. FACT: fine-grained cross-media interaction with documents via a portable hybrid paper-laptop interface. In *Proceedings of the international conference on Multimedia - MM '10*, MM '10, pages 361–370, New York, New York, USA, 2010. ACM Press. ISBN 9781605589336. doi: 10.1145/1873951.1874001. URL http://doi.acm.org/10.1145/1873951.1874001. (Cited on pages 121 and 145.)

Roman Lissermann, Simon Olberding, Max Mühlhäuser, and Jürgen Steimle. Interacting with videos on paper-like displays. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*, CHI EA '12, pages 2579–2584, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1016-1. doi: 10.1145/2212776.2223839.

URL http://doi.acm.org/10.1145/2212776.2223839. (Cited on page 155.)

Markus Löchtefeld, Johannes Schöning, Michael Rohs, and Antonio Krüger. Marauders light: replacing the wand with a mobile camera projector unit. In *Mobile and Ubiquitous Multimedia*, 2009. URL http://portal.acm.org/citation.cfm?id=1658550.1658569. (Cited on page 122.)

Julian Looser, Mark Billinghurst, and Andy Cockburn. Through the looking glass: the use of lenses as an interface tool for augmented reality interfaces. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, GRAPHITE '04, pages 204–211, New York, NY, USA, 2004. ACM. ISBN 1-58113-883-0. doi: 10.1145/988834.988870. URL http://doi.acm.org/10.1145/988834.988870. (Cited on page 81.)

David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004. ISSN 0920-5691. doi: 10.1023/B:VISI.0000029664.99615.94. URL http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94. (Cited on pages 143 and 145.)

Kris Luyten, Kristof Verpoorten, and Karin Conix. Ad-hoc co-located collaborative work with mobile devices. In *Proc. mobileHCI '07*, pages 507–514. ACM New York, NY, USA, 2007. (Cited on page 82.)

I. Scott MacKenzie and William Buxton. Extending fitts' law to two-dimensional tasks. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 219–226, New York, NY, USA, 1992. ACM. (Cited on page 100.)

Charlotte Magnusson, Annika Waern, Kirsten Rassmus Gröhn, Ase Bjernryd, Helen Bernhardsson, Ann Jakobsson, Johan Salo, Magnus Wallon, and Per-Olof Hedvall. Navigating the world and learning to like it: mobility training through a pervasive game. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 285–294, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0541-9. doi: 10.1145/2037373.2037416. URL http://doi.acm.org/10.1145/2037373.2037416. (Cited on page 80.)

M. E. J. Masson, W. C. Hill, J. Conner, and R. Guidon. Misconceived misconceptions? In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '88, pages 151–156, New York, NY, USA, 1988. ACM. ISBN 0-201-14237-6. doi: http://doi.acm.org/10.1145/57167.57192. URL http://doi.acm.org/10.1145/57167.57192. (Cited on page 65.)

Sumit Mehra, Peter Werkhoven, and Marcel Worring. Navigating on handheld displays: Dynamic versus static peephole navigation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(4), 2006. ISSN 1073-0516. URL http://portal.acm.org/citation.cfm?id=1188816.1188818. (Cited on page 80.)

Paul Milgram and Fumio Kishino. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, E77-D(12), December 1994. (Cited on pages 78, 79 and 182.)

David Molyneaux and Hans Gellersen. Projected interfaces. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction - TEI '09*, page 385, New York, New York, USA, February 2009. ACM Press. ISBN 9781605584935. doi: 10.1145/1517664.1517741. URL http://dl.acm.org/citation.cfm?id=1517664.1517741. (Cited on page 123.)

David Molyneaux, Hans Gellersen, Bernt Schiele, Daniel Roggen, Clemens Lombriser, Gerhard Tröster, Gerd Kortuem, and Paul Havinga. *Smart Sensing and Context*, volume 5279 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-88792-8. doi: 10.1007/978-3-540-88793-5. URL http://www.springerlink.com/content/ut74k7258110733w/. (Cited on page 123.)

David Molyneaux, Shahram Izadi, David Kim, Otmar Hilliges, Steve Hodges, Xiang Cao, Alex Butler, and Hans Gellersen. Interactive environment-aware handheld projectors for pervasive computing spaces. In *Proceedings of the 10th international conference on Pervasive Computing*, Pervasive '12, pages 197–215, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-31204-5. doi: 10.1007/978-3-642-31205-2_13. URL http://dx.doi.org/10.1007/978-3-642-31205-2_13. (Cited on page 122.)

Sugata Mukhopadhyay and Brian Smith. Passive capture and structuring of lectures. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, MULTIMEDIA '99, pages 477–487, New York, NY, USA, 1999. ACM. ISBN 1-58113-151-8. doi: 10.1145/319463.319690. URL http://doi.acm.org/10.1145/319463.319690. (Cited on page 22.)

Michael Naimark. Two unusual projection spaces. *Presence: Teleoper. Virtual Environ.*, 14(5):597–605, October 2005. ISSN 1054-7460. URL http://dl.acm.org/citation.cfm?id=1160382.1160391. (Cited on page 123.)

Donald Norman. *The Design of Everyday Things*. Perseus Books, 2002. (Cited on pages 66 and 156.)

Donald A. Norman and Stephen W. Draper. *User Centered System Design; New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1986. ISBN 0898597811. (Cited on page 12.)

Kenton O'Hara, April Slayden Mitchell, and Alex Vorbau. Consuming video on mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 857–866, San Jose, California, USA, 2007. ACM. (Cited on pages 16 and 17.)

Thomas Pederson and Dipak Surie. Towards an activity-aware wearable computing platform based on an egocentric interaction model. In *UCS '07*, pages 211–227, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 3-540-76771-1, 978-3-540-76771-8. (Cited on page 93.)

Martin Pielot, Benjamin Poppinga, Wilko Heuten, and Susanne Boll. A tactile compass for eyes-free pedestrian navigation. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part II*, INTERACT'11, pages 640–656, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-23770-6. URL http://dl.acm.org/citation.cfm?id=2042118.2042179. (Cited on page 80.)

Gonzalo Ramos and Ravin Balakrishnan. Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, UIST '03, pages

105–114, New York, NY, USA, 2003. ACM. ISBN 1-58113-636-6. doi: http://doi.acm.org/10.1145/964696.964708. (Cited on pages 21 and 25.)

Ramesh Raskar, Paul Beardsley, Jeroen van Baar, Yao Wang, Paul Dietz, Johnny Lee, Darren Leigh, and Thomas Willwacher. RFIG lamps: interacting with a self-describing world via photosensing wireless tags and projectors. In *International Conference on Computer Graphics and Interactive Techniques*, volume 23, 2004. URL http://portal.acm.org/citation.cfm?id=1186562.1015738. (Cited on page 122.)

Ramesh Raskar, Jeroen van Baar, Paul Beardsley, Thomas Willwacher, Srinivas Rao, and Clifton Forlines. iLamps: geometrically aware and self-configuring projectors. In *International Conference on Computer Graphics and Interactive Techniques*, 2006. URL http://portal.acm.org/citation.cfm?id=1185657.1185802. (Cited on page 122.)

Maurizio Rigamonti, Denis Lalanne, and Rolf Ingold. Faericworld: browsing multimedia events through static documents and links. In *Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction*, INTERACT'07, pages 102–115, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 3-540-74794-X, 978-3-540-74794-9. URL http://dl.acm.org/citation.cfm?id=1776994.1777009. (Cited on page 44.)

Simon Robinson, Parisa Eslambolchilar, and Matt Jones. Sweep-shake: finding digital resources in physical environments. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 12:1–12:10, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-281-8. doi: 10.1145/1613858.1613874. URL http://doi.acm.org/10.1145/1613858.1613874. (Cited on page 79.)

Yvonne Rogers. Hci theory: Classical, modern, and contemporary. *Synthesis Lectures on Human-Centered Informatics*, 5(2):1–129, 2012. doi: 10.2200/S00418ED1V01Y201205HCI014. URL http://www.morganclaypool.com/doi/abs/10.2200/S00418ED1V01Y201205HCI014. (Cited on page 12.)

Michael Rohs and Georg Essl. Which one is better?: information navigation techniques for spatially aware handheld displays. In *International Con-*

ference on Multimodal Interfaces, 2006. URL http://portal.acm.org/citation.cfm?id=1180995.1181016. (Cited on page 81.)

Michael Rohs and Georg Essl. Sensing-based interaction for information navigation on handheld displays. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pages 387–394, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-862-6. doi: http://doi.acm.org/10.1145/1377999.1378043. (Cited on page 81.)

Michael Rohs and Antti Oulasvirta. Target acquisition with camera phones when used as magic lenses. In *Conference on Human Factors in Computing Systems*, 2008. URL http://portal.acm.org/citation.cfm?id=1357054.1357275. (Cited on pages 81, 99, 100 and 101.)

Michael Rohs, Georg Essl, Johannes Schöning, Anja Naumann, Robert Schlecher, and Antonio Krüger. Impact of item density on magic lens interactions. In *ACM International Conference Proceeding Series*, 2009. URL http://portal.acm.org/citation.cfm?id=1613858.1613907. (Cited on page 81.)

Michael Rohs, Antti Oulasvirta, and Tiia Suomalainen. Interaction with magic lenses: real-world validation of a fitts' law model. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2725–2728, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979343. URL http://doi.acm.org/10.1145/1978942.1979343. (Cited on page 101.)

Lawrence A. Rowe, Diane Harley, Peter Pletcher, and Shannon Lawrence. Bibs: A lecture webcasting system. Technical report, Center for Studies in Higher Education University of California, 2001. (Cited on page 22.)

Enrico Rukzio, Paul Holleis, and Hans Gellersen. Personal Projectors for Pervasive Computing. *IEEE Pervasive Computing*, February 2011. ISSN 1536-1268. doi: 10.1109/MPRV.2011.17. URL http://www.computer.org/portal/web/csdl/doi/10.1109/MPRV.2011.17. (Cited on page 119.)

Dieter Schmalstieg, Anton Fuhrmann, Gerd Hesina, Zsolt Szalavári, L. Miguel Encarnaçäo, Michael Gervautz, and Werner Purgathofer. The studierstube augmented reality project. *Presence: Teleoper. Virtual Environ.*, 11(1):33–54, February 2002. ISSN 1054-7460. doi: 10.1162/105474602317343640. URL `http://dx.doi.org/10.1162/105474602317343640`. (Cited on page 81.)

Klaus Schoeffmann and Manfred del Fabro. Hierarchical video browsing with a 3d carousel. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 827–828, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0616-4. doi: 10.1145/2072298.2072479. URL `http://doi.acm.org/10.1145/2072298.2072479`. (Cited on page 24.)

Klaus Schoeffmann, Mario Taschwer, and Laszlo Boeszoermenyi. The video explorer: a tool for navigation and searching within a single video based on fast content analysis. In *Proceedings of the rst annual ACM SIGMM conference on Multimedia systems*, MMSys '10, pages 247–258, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-914-5. doi: 10.1145/1730836.1730867. URL `http://doi.acm.org/10.1145/1730836.1730867`. (Cited on page 22.)

Johannes Schöning, Michael Rohs, Sven Kratz, Markus Löchtefeld, and Antonio Krüger. Map torchlight: a mobile augmented reality camera projector unit. In *Conference on Human Factors in Computing Systems*, 2009. URL `http://portal.acm.org/citation.cfm?id=1520581`. (Cited on page 122.)

Tim Schwanen. Urban form and commuting behaviour: a crosseuropean perspective. *Tijdschrift voor Economische en Sociale Geogra e, Royal Dutch Geographical Society KNAG*, 93(3):336–343, 2008. (Cited on page 18.)

Helen Sharp, Yvonne Rogers, and Jenny Preece. *Interaction Design: Beyond Human-Computer Interaction*. Wiley, 2 edition, March 2007. ISBN 0470018666. URL `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0470018666`. (Cited on page 11.)

Frank Shipman, Andreas Girgensohn, and Lynn Wilcox. Authoring, viewing, and generating hypervideo: An overview of Hyper-Hitchcock. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOM-*

CCAP), 5(2), 2008. ISSN 1551-6857. URL `http://portal.acm.org/citation.cfm?id=1413862.1413868`. (Cited on page 24.)

Hyunyoung Song, Tovi Grossman, George Fitzmaurice, François Guimbretiere, Azam Khan, Ramtin Attar, and Gordon Kurtenbach. PenLight: combining a mobile projector and a digital pen for dynamic visual overlay. In *Conference on Human Factors in Computing Systems*, 2009. URL `http://portal.acm.org/citation.cfm?id=1518701.1518726`. (Cited on page 122.)

Hyunyoung Song, Francois Guimbretiere, Tovi Grossman, and George Fitzmaurice. MouseLight: bimanual interactions on digital paper using a pen and a spatially-aware mobile projector. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, CHI '10, pages 2451–2460, New York, New York, USA, 2010. ACM Press. ISBN 9781605589299. doi: 10.1145/1753326.1753697. URL `http://doi.acm.org/10.1145/1753326.1753697`. (Cited on page 122.)

Martin Spindler, Christian Tominski, Heidrun Schumann, and Raimund Dachselt. Tangible views for information visualization. In *ACM International Conference on Interactive Tabletops and Surfaces - ITS '10*, pages 157–166, New York, New York, USA, November 2010. ACM Press. ISBN 9781450303996. doi: 10.1145/1936652.1936684. URL `http://dl.acm.org/citation.cfm?id=1936652.1936684`. (Cited on pages 82 and 132.)

A. Strauss and J. Corbin. Basics of qualitative research: Grounded theory procedures and techniques. 1990. (Cited on page 86.)

A Strauss and J Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, volume 2nd. Sage Publications, 2008. ISBN 0803959400. (Cited on pages 11, 86, 88 and 127.)

Lucy A. Suchman. *Plans and situated actions - the problem of human-machine communication*. Learning in doing: social,cognitive,and computational perspectives. Cambridge University Press, 1987. ISBN 978-0-521-33739-7. (Cited on page 11.)

Qibin Sun and W. Hurst. Video browsing on handheld devices–interface designs for the next generation of mobile video players. *Multimedia, IEEE*, 15(3):76

–83, july-sept. 2008. ISSN 1070-986X. doi: 10.1109/MMUL.2008.66. (Cited on pages 17, 26 and 45.)

Ba Tu Truong and Svetha Venkatesh. Video abstraction: A systematic review and classification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3, February 2007. ISSN 1551-6857. doi: http://doi.acm.org/10.1145/1198302.1198305. URL `http://doi.acm.org/10.1145/1198302.1198305`. (Cited on page 44.)

H.D. Wactlar, T. Kanade, M.A. Smith, and S.M. Stevens. Intelligent access to digital video: Informedia project. *Computer*, 29(5):46–52, 1996. (Cited on page 24.)

Daniel Wagner and Dieter Schmalstieg. Handheld augmented reality displays. *Virtual Reality Conference, 2006*, 2006. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1667684`. (Cited on page 78.)

Karl D.D. Willis, Ivan Poupyrev, Scott E. Hudson, and Moshe Mahler. Sidebyside: ad-hoc multi-user interaction with handheld projectors. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 431–440, New York, NY, USA, 2011a. ACM. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047254. URL `http://doi.acm.org/10.1145/2047196.2047254`. (Cited on page 122.)

Karl D.D. Willis, Ivan Poupyrev, and Takaaki Shiratori. Motionbeam: a metaphor for character interaction with handheld projectors. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, pages 1031–1040, New York, New York, USA, May 2011b. ACM Press. ISBN 9781450302289. doi: 10.1145/1978942.1979096. URL `http://dl.acm.org/citation.cfm?id=1978942.1979096`. (Cited on page 122.)

Andrew D. Wilson and Hrvoje Benko. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pages 273–282, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0271-5. doi: 10.1145/1866029.1866073. URL `http://doi.acm.org/10.1145/1866029.1866073`. (Cited on page 123.)

Max L. Wilson, Simon Robinson, Dan Craggs, Kristian Brimble, and Matt Jones. Pico-ing into the future of mobile projector phones. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*, CHI EA '10, pages 3997–4002, New York, New York, USA, 2010. ACM Press. ISBN 9781605589305. doi: 10.1145/1753846.1754092. URL `http://doi.acm.org/10.1145/1753846.1754092`. (Cited on page 122.)

Zi Ye and Hammad Khalid. Cobra: flexible displays for mobilegaming scenarios. *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*, pages 4363–4368, 2010. doi: 10.1145/1753846.1754154. URL `http://portal.acm.org/citation.cfm?doid=1753846.1754154`. (Cited on page 123.)

Ka-Ping Yee. Peephole displays: pen interaction on spatially aware handheld computers. In *Proc. CHI '03*, pages 1–8, New York, NY, USA, 2003. ACM. ISBN 1-58113-630-7. doi: http://doi.acm.org/10.1145/642611.642613. (Cited on pages 80, 81 and 182.)

YouTube. Press statistics (last checked: October 29, 2012), 2012. URL `http://www.youtube.com/t/press_statistics/`. (Cited on page 1.)

Cha Zhang, Yong Rui, Jim Crawford, and Li-Wei He. An automated end-to-end lecture capture and broadcasting system. *ACM Trans. Multimedia Comput. Commun. Appl.*, 4(1):6:1–6:23, February 2008. ISSN 1551-6857. doi: 10.1145/1324287.1324293. URL `http://doi.acm.org/10.1145/1324287.1324293`. (Cited on page 22.)

Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '08, pages 193–202, Washington, DC, USA, 2008. IEEE Computer Society. ISBN 978-1-4244-2840-3. doi: 10.1109/ISMAR.2008.4637362. URL `http://dx.doi.org/10.1109/ISMAR.2008.4637362`. (Cited on page 78.)

List of Figures

List of Tables

Wissenschaftlicher Werdegang des Autors

2003-2009	Studium der Informatik an der Technischen Universität Darmstadt Abschluss: *Diplom Informatiker*
2005-2008	Studium der Mathematik an der Technischen Universität Darmstadt Abschluss: *Diplom Mathematiker*
2008-2012	Stipendiat im DFG Graduiertenkolleg E-Learning (GRK-1223) Wissenschaftlicher Mitarbeiter am Lehrstuhl Telekooperation, Fachbereich Informatik, an der Technischen Universität Darmstadt